BMJ Health & Care Informatics

Involving multiple stakeholders in assessing and reviewing a novel data visualisation tool for a national neonatal data asset

William Bishop Lammons ,[1,2] Becky Moss,[2] Charlie Bignell,[2] Chris Gale,[2] Adam MacBride,[3] Ricardo Ribas,[2] Cheryl Battersby,[2] Neena Modi[2]

To cite: Lammons WB, Moss B, Bignell C, *et al*. Involving multiple stakeholders in assessing and reviewing a novel data visualisation tool for a national neonatal data asset. *BMJ Health Care Inform* 2023;**30**:e100694. doi:10.1136/bmjhci-2022-100694

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjhci-2022-100694).

Received 24 October 2022
Accepted 07 January 2023

[1]Department of Applied Health Research, UCL, London, UK
[2]Section of Neonatal Medicine, School of Primary Care and Public Health, Imperial College London, London, UK
[3]Public co-author, n/a, London, UK

Correspondence to
William Bishop Lammons; william.lammons@ucl.ac.uk

ABSTRACT

Objectives We involved public and professional stakeholders to assess a novel data interrogation tool, the Neonatal Health Intelligence Tool, for a National Data Asset, the National Neonatal Research Database.

Methods We recruited parents, preterm adults, data managers, clinicians, network managers and researchers (trialists and epidemiologists) for consultations demonstrating a prototype tool and semi-structured discussion. A thematic analysis of consultations is reported by stakeholder group.

Results We held nine on-line consultations (March–December 2021), with 24 stakeholders: parents (n=8), preterm adults (n=2), data managers (n=3), clinicians (n=3), network managers (n=2), triallists (n=3) and epidemiologists (n=3). We identified four themes from parents/preterm adults: struggling to consume information, Dads and data, bring data to life and yearning for predictions; five themes from data managers/clinicians/network managers: benchmarking, clinical outcomes, transfers and activity, the impact of socioeconomic background and ethnicity, and timeliness of updates and widening availability; and one theme from researchers: interrogating the data.

Discussion Other patient and public involvement (PPI) studies have reported that data tools generate concerns; our stakeholders had none. They were unanimously supportive and enthusiastic, citing visualisation as the tool's greatest strength. Stakeholders had no criticisms; instead, they recognised the tool's potential and wanted more features. Parents saw the tool as an opportunity to inform themselves without burdening clinicians, while clinicians welcomed an aid to explaining potential outcomes to parents.

Conclusion All stakeholder groups recognised the need for the tool, praising its content and format. PPI consultations with all key groups, and their synthesis, illustrated desire for additional uses from it.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Patient and public involvement provides benefits across all aspects of health and social care research and should be the gold standard from study inception to dissemination and evaluation.
⇒ Data interrogation tools allow for analysis of large-scale point-of-care data on a nationwide scale.

WHAT THIS STUDY ADDS

⇒ This study demonstrates the importance of extending standard patient and public involvement to include a wider range of stakeholders and synthesising the contributions of each group to maximise the uses and value of data interrogation tools.
⇒ This study illustrates how qualitative analysis can identify connections across differing stakeholder groups in the context of a public involvement activity.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ An effective and useful data interrogation tool will encourage national and international researchers to use the health data contained in the National Neonatal Research Database.
⇒ Clinicians will be able to use the tool in their practice as an aid to explaining neonatal outcomes to parents.
⇒ Interrogation of the data set will enable policy makers to explore influences on neonatal outcomes.

INTRODUCTION

The National Neonatal Research Database

The National Neonatal Research Database (NNRD) is a data asset containing detailed clinical information, a standard extract from the Electronic Patient Records of admissions to all NHS neonatal units in England, Wales, Scotland and the Isle of Man.[1 2] The extract (the Neonatal Data Set, an approved NHS Information Standard) undergoes quality assurance prior to deposition in the NNRD. The NNRD is a UK Research Ethics Service–approved database (10/80803/151) and supports national and international neonatal research. Access is through the Health Data Research UK Gateway (https://www.imperial.ac.uk/neonatal-data-analysis-unit/neonatal-data-analysis-unit/utilising-the-national-neonatal-research-database/).

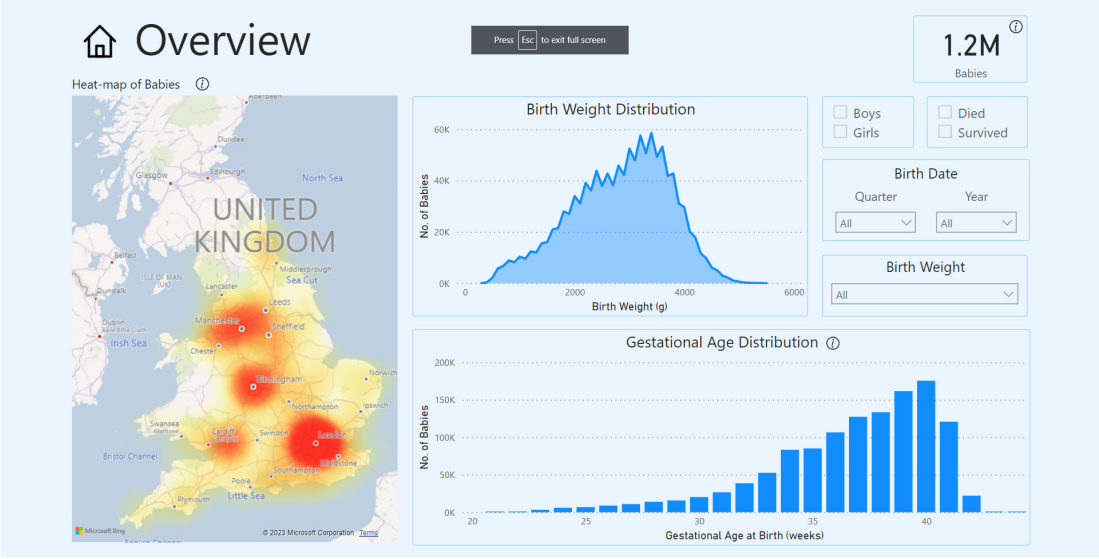

Figure 1 Neonatal Health Intelligence Tool, Overview feature showing birth weight, gestational age trends across the UK.

In 2020, with the support from Medical Research Council, we began developing a Neonatal Health Intelligence Tool (figures 1 and 2; https://www.imperial.ac.uk/neonatal-data-analysis-unit/neonatal-data-analysis-unit/neonatal-data-visualisations/) that enables interrogation of the NNRD[3] and viewing of data on babies requiring neonatal care from 2008 by neonatal network. The tool is an online web-based application using graphs and charts that show data and trends on neonatal outcomes, such as necrotising enterocolitis and bronchopulmonary dysplasia, in a visual, easy-to-use format.

Parent, patient and public involvement

Commonly termed 'Patient and Public Involvement' (PPI), neonatal medicine uses 'Parent, Patient and Public Involvement' (PPPI), which focuses on collaborating with parents, patients and members of the public to ensure those impacted by studies contribute to their design and delivery.[4–6] Some PPPI work combines the voices of parents, patients and the public with those of clinical or research staff,[7–9] but these consultations are commonly reported discretely by individual groups: the public, clinicians or researchers.[10 11] These consultations' outputs are also typically patient-focused communication materials or decision-making tools, rather than data visualisation tools.

National Institute of Health and Care Research (NIHR) standards of public involvement focus on patients and the public as service users, without considering clinician or researcher stakeholder roles.[12] However, since the NNRD data tool benefits a wide range of stakeholders, we applied the concepts and methods of public and patient consultations to an exhaustive list of professional stakeholders

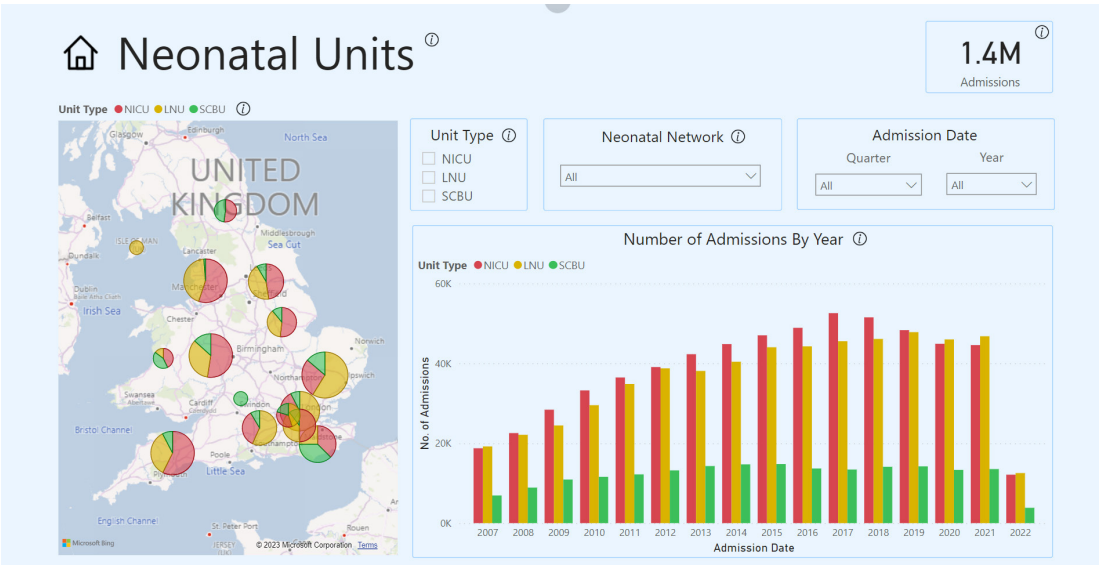

Figure 2 Neonatal Health Intelligence Tool, Survival feature showing rates of survival against birth weight and gestational age variables.

who would use these tools to explore neonatal activity and outcomes. We therefore term this approach 'public and stakeholder involvement,' a technique which has received support[9] but has been applied to few data visualisation tools.

Study objectives

Our objectives for these stakeholder consultations were to:

- Understand what all stakeholder groups need from the tool
- Identify functionality issues
- Optimise the efficacy of the design process by synthesising all stakeholder views

METHODS

Recruitment

We recruited parents and adults who were born preterm ('preterm adults') through an existing network of individuals (neoWONDER; https://www.neowonder.org.uk/) who consented to be invited to participate in research by email,[13] and professionals through our contacts and NHS network lists, also by email.

Consultation structure

We combined PPPI with qualitative methodologies to learn about each stakeholder group's unique needs.[14 15] Each consultation was attended by one to four participants; parent and preterm adults were mixed groups, and the remaining stakeholders were recruited to attend groups based on their specific requirements of the NNRD. Each session featured:

1. A demonstration of the NNRD Neonatal Health Intelligence Tool's features led by their designer (CB), lasting approximately 15 min (figures 1 and 2)
2. Stakeholder question-and-answer session in response to the demonstration.
3. Semi-structured focus group discussion (individual interviews, n=2, used the same topic guide)

Sessions were held by video Zoom meeting. Participant video was on by default, but to support inclusivity, participants were encouraged to have video on or off at their discretion. Session lengths averaged 75 min and were video recorded with verbal consent. Zoom's closed captioning function was used to create a raw transcript which was saved and then edited by the authors for clarity. The focus group topic guide incorporated deductive questions (eg, '*would this be useful?*') and inductive questions (eg, *'why would this be useful?'*), to allow further ideas to emerge from discussion.[15] A complete list of topic guide questions is included in online supplemental appendix 1. The semi-structured format allowed flexibility to probe topics particular to the discussion and the stakeholder audience.[16] The social context of sharing and discussing ideas in a focus group format also allowed for 'reflection and refinement' of participants' ideas, which can "deepen respondents' insights into their own circumstances, attitudes or behaviour".[17] We offered participants the chance to contribute through the chat function, speak in an individual interview and called participants individually during meetings to ensure all had a chance to verbalise their opinions. We stopped recruiting when we ceased to obtain further insights; this is referred to in qualitative methodology as data saturation.[18] Professional consultations were chaired by a neonatologist (CG) and parent/preterm adult consultations by the PPPI Leads (BM, WL).

Analysis

We combine approaches for understanding relationships between users and digital tools, reporting thematic analysis of consultations by stakeholder group.[17 19] BM and WBL analysed transcripts by individual stakeholder group using each using a combination of NVivo V.1.3, NVivo V.16 (QSR International) and Microsoft Word. This revealed the unique needs of each, while highlighting commonalities between them. We reviewed transcripts on an individual stakeholder group basis, then created thematic codes to identify key subthemes.[20] WL then conducted a final iterative analysis comparing and synthesising themes across differing stakeholder groups. BM and WBL reviewed one another's analyses throughout this process to ensure themes were comprehensively captured.[21] We report key PPI checklist items in the Guidance for Reporting Involvement of Patients and the Public 2 (GRIPP2, online supplemental appendix 2).[22]

RESULTS

We held nine digital consultation groups between March and December 2021, with 24 stakeholders: parents (n=8), preterm adults (n=2), data managers (n=3), clinicians (n=3), network managers (n=2), triallists (n=3), and epidemiologists (n=3).

Parents and preterm adults

The tool was enthusiastically received by parents (Mums=6, Dads=2) and preterm adults and described as something they would use. The visual nature was regarded as helpful and easy to grasp. Four salient themes were identified: struggling to consume information, Dads and data, bringing the data to life and yearning for predictions.

Struggling to consume information

Parents described difficulty with comprehending health information due to the emotional strain of their baby being on the neonatal unit. Indeed, one felt the amount of data featured in the tool was excessive:

> '*That is way too much data… I think I would sit down… look at that… and say, "Oh my god", and… panic'* (Mum).

However, while another participant concurred that the tool contained a large amount of information, they indicated that some parents desired even more than this. They described a complex balance between the volume of data required for decision-making and the type or amount of

information capable of worsening their distress, particularly when the baby's survival is precarious:

> *'…We had [him at] 23 [weeks]… [the doctors] said: "Baby is going to be born. We can either take them away from you for… assessment, but by the time we hand him back to you, he might already be dead. Or you can give him a cuddle as he passes away. We'll come back in 30 minutes…" It is horrific. A decision needs to be made, ultimately, so you need… to know these things…'* (Mum).

Dads and data

Across all parent and preterm adult consultation groups, fathers expressed interest in statistics on neonatal conditions and complications, saying: 'I wanted the stats,' or: 'I wanted numbers.' They described a gendered division of parental labour, with mothers handling tasks 'closer' to the baby, such as expressing milk. These fathers regarded affinity for data as supporting their baby and combating their inability to fulfil stereotypical masculine ideals:

> *'They were telling us to stop looking at the machine and look at the baby. I couldn't help [it]. I stood with the machine, I was analysing numbers… As a man, you want to try and protect them. You want… to try and do something, and it's completely useless'* (Dad).

Bringing the data to life

Several participants desired to have individual case stories that represented their experiences; for them, the scale of the data muted how they related to the experiences of other families:

> *'You really need to provide actual stories of patients and parents, so that those statistics are not [just] statistics. So when you click on some of those dots, they take you to parent voices, interviews – good and bad'* (Preterm Adult Woman).

Yearning for predictions

Many participants expressed a need for, and limited access to, clear statements explaining risks, in a format such as: 'X out of 10 000 preterm babies will experience this.' They recognised each baby's outcomes would be 'different', complicating prediction. However, they believed accessing large-scale data visualisations could help alleviate this:

> *'I asked for kind of stats because I needed to prepare myself… I had four days… so I think being armed with some form of stats [would have been helpful] '* (Dad).

Although the tools were not designed to predict outcomes, one parent perceived data—specifically probability data—as being particularly helpful in navigating the swiftly changing risks that neonatal babies face:

> *'You think "we're breezing this"… then it really hits the fan… [data] gives you something to hold onto… it allows you to manage your expectations… [the consultants'] knowledge isn't going to be anywhere near as broad as the data… the conversation I had with them all the time was, " what are the chances of X?" " I don't know, all babies are different." "Okay but if you've got 10 k of these babies, what are the chances?" "We haven't got that information"'* (Dad).

Parents pointed out that accessing the tool would offer an opportunity to reduce encroaching on clinicians' time with requests for detailed explanations.

Data managers, clinicians and network managers

The data managers, clinicians and network managers were recruited from the same neonatal network, and some regularly worked closely; therefore, the five themes which emerged from their consultations are reported together. The themes were comparativeness of data across units, additional outcomes and transfers, socioeconomic status and ethnicity, timeliness of data updates and widening availability of the tool.

Comparativeness of data across units

The most salient theme for this group was to understand differences between their unit's outcomes and those of other comparable units:

> *'Something that allows us to make sure that we're comparing like for like and that we're not comparing a cohort of babies in… [our network] and thinking we're doing terribly because [another unit is] doing much better… [when] actually their data is different…'* (Data Manager).

Presently, units define their standards by routinely monitoring data and comparing themselves to trends set by the collective work of units, but this can be challenging:

> *'I can pull things out from [electronic patient record system] but it might be very complicated… Rather than me having to constantly monitor data… if every year, I'm going to look at these lines illustrating differences across units, I know where I am heading to and where I need to work'* (Data Manager).
>
> *'Allowing us to identify where those differences in practice may be occurring, and pointers as to where we can go to either to review ourselves compared to peers around the country or which units we need to… focus on'* (Data Manager).

This group also focused heavily on the opportunities the tool would offer such as direct comparisons with other units. They recommended these be made between overarching operational delivery networks (ODNs) and between individual units within an ODN. One network manager reported that they were happy to feed back to their wider network, and expressed interest in sharing public-facing network-level data if anonymised and password protected:

> *'…I can see this taking off, it's very detailed and it provides a lot of flexibility… there's room for it to grow and develop… there's lots of things like audit and parent engagement, I can imagine they'd like to see this data to kind of reassure them; it covers all the bases'* (Network Manager).

One network manager discussed specific ODN needs:

'It will certainly aid the standardisation of review… the ODNs all have a data manager, but there's variance… each staff member is very different, and the needs of networks are quite different: it's a genuine comparison with like for like… a huge step forward' (Network Manager).

Another agreed, saying:

'Benchmarking is one of the key things we've struggled with for such a long time. Locally, we can compare apples with apples but my data and other managers' would be apples and pears. That's where the tools come into their own [sic]' (Network Manager).

Additional outcomes and transfers

Discussion of additional outcomes beyond survival and length of stay focused on understanding reasons for admission and its variation between units. However, participants also expressed the desire for data on a wide range of outcomes, including breast feeding, intraventricular haemorrhage, modifiable factors, tube feeding in hospital and on discharge, probiotics, necrotising enterocolitis, lung disease, ventilator use and transfers. Transfers were also discussed as a particularly difficult phenomenon to track because of the varying reasons for why they occur:

'…I know there are two units whose babies should be coming to me – if they're not, why? We don't capture that… so we have to rely on audits…. It's a big exercise capturing that, and the focus nationally is on trying to reduce transfers. What is their trajectory?' (Clinician).

Socioeconomic status and ethnicity

Central to comparing units and ODN was identifying areas of disproportionate need, particularly how socioeconomic status and ethnicity shape their neonatal unit populations and admissions:

'…Some of the sliders in the tool would filter and say, "Out of your population of your 24–26 weekers, view that by ethnicity or deprivation"… It can help us understand where the maternity drivers need to be and to help some of those mums in a more focused and targeted way to understand getting better antenatal care…' (Clinician).

Timeliness of data updates

Updating neonatal data was an integral complexity, caused by the length of time before results of clinical practice adjustments were observable:

'If we made a change, we're not going to see it for another year…. If we've been doing the wrong thing, it takes us a long time to find out' (Data Manager).

Participants said the visualisation tool could provide 'standardisation' of and 'more frequent access to' data on a regular basis, but neonatal care processes cause a 'lag' in data access:

'Anything we do monthly we're not going to have any outcomes unless they died within a few days. We don't know lengths of stay for 3 or 4 months in some cases… Having it too quickly is detrimental, we don't get the true picture…' (Data Manager).

One participant emphasised the solution would lie in 'access to timely data but not too timely data.'

Widening availability of the tool

One clinician described themself as a 'technical dinosaur' and, like the parent and preterm adult groups, said the visual nature of the tool was helpful. They also emphasised the importance of sharing visualisations with parents as PDF printouts and expressed strong interest in distributing the tool demonstration to data manager groups. They added that the inclusion of explanatory notes or supplementary material within the tool instead of simplifying existing content could support future interpretation. The network managers did not anticipate parental concerns regarding data use:

'…They trust that professionals are doing that judgement call. If it's good valid data, that's why they're showing them; as opposed to parents questioning data, then questioning what the answers are…' (Network Manager).

The only exception was where there were very low numbers, for example, rare congenital anomalies, which could, coupled with geographical location or gestational age, identify babies. They felt confident they could reassure parents that data were insufficiently fine-grained for this.

Researchers (trialists and epidemiologists)

A single theme emerged from the consultations with researchers: interrogating the data in the tool.

Interrogating the data

Trialists and epidemiologists enquired about the tool's versatility, praising its usefulness for grants and study planning. They asked whether data will be broken down by conditions, and if capability exists for cross tabulations of more than two variables, such as filters exploring birth weight, gestational age and status 1 year later. One said:

'That would be really helpful for planning any study, not just trials, to get down to that level' (Epidemiologist).

Trialists, like network managers, advocated for supplementary information including a data dictionary, suggesting this could be 'hover and click.' Like data managers, they cautioned against fine-grained attributes, explaining that, for example, the small number of 500 g birthweight babies 'etch themselves on your brain.' Another triallist expressed desire for data categories, 'given the data spans ten years' (Triallist).

Another broadened this point, emphasising the inclusion of missing data in categories:

> *'It is still useful to have categories where there is a lot of missingness, even if they are not usable in a trial, as an indication of what is not currently routinely collected'* (Triallist).

DISCUSSION

We conducted consultations with 24 individuals from seven different stakeholder groups: parents, preterm adults, data managers, clinicians, network managers, trialists and epidemiologists. We identified four themes from parents and preterm adults: struggling to consume information, Dads and data, bringing the data to life and yearning for predictions; five themes from data managers, clinicians and network managers: comparativeness of data across units, additional outcomes and transfers, socioeconomic background and ethnicity, timeliness of data updates and widening the tool's availability; and one theme from trialists and epidemiologists: interrogating the data.

PPPI on health data typically focuses on artificial intelligence,[23] patient portals[24] and decision-making tools[25] and recommend problem solving by understanding the relationship between data capabilities and the role of patients and the public.[23–26] Few involvement activities have focused specifically on providing public-facing data visualisations that are also useful to researchers, data professionals and clinicians. Our work contributes to the literature by consulting with all relevant stakeholder groups and synthesising their contributions. This enabled us to create and refine a tool which is useful to all groups, for a range of purposes specific to each. While other studies report that an artificial intelligence tool generated concerns,[23] our stakeholders reported no concerns but were unanimously supportive and enthusiastic, citing visualisation functions as its greatest strength. Stakeholders had almost no criticisms of the tool, instead recognising its potential and desiring additional features. Parents saw the tool as an opportunity to inform themselves without burdening clinical staff with questions, while clinicians welcomed data visualisation as an aid to explaining potential outcomes to parents. All stakeholder groups engaged strongly with the tool's potential and sought additional features. This suggests a universal need among the public, research and academic communities for more ways to address the broad, challenging and anxiety-inducing uncertainties in neonatal care.

The strength of this study was the inclusion of all seven relevant stakeholder groups and the very high level of consensus they reached. Study limitations are that participants were a self-selecting convenience sample, all were English speaking and a large group came from the same network, which may mean they are not representative of the wider population of stakeholders. Other limitations included the disproportionate gender balance in favour of mothers in the parent and preterm adult focus groups, and the potential for the limited number of fathers to have disproportionately affected the results. However, given that there is an underrepresentation of fathers in research involvement, any findings that such a consultation can generate are helpful.

We recommend that future health data involvement and engagement work should involve all stakeholders to form a full picture of perspectives and requirements. We intend to investigate the possibility of embedding parent and preterm adult narratives in the Neonatal Health Intelligence Tool alongside the statistical data, to enrich and humanise the experiences embodied by the data.

CONCLUSIONS

All consultation participants united behind a need for the tool, praised its content and format, and desired additional features. Our consultations and qualitative inquiry with all groups of key stakeholders has enabled us to ensure the tool's relevance and value to the communities it serves.

Twitter William Bishop Lammons @william_lammons

Acknowledgements We would like to thank all parents, adults born preterm, data professionals, clinicians and researchers who participated in this public and stakeholder involvement consultation for the NNRD Neonatal Health Intelligence Tool.

Contributors WBL: writing, draft editing, thematic analysis, qualitative design, public involvement design data collection, consultation moderation. BM: writing, draft editing, thematic analysis, qualitative design, public involvement design, data collection, consultation moderation. ChaB: tool design, tool demonstration, public engagement tool Q&A, tool graphics creation. CG: consultation moderation, writing, draft editing. AM: writing, draft editing. RR: writing, draft editing. CheB: draft editing, public involvement participant recruitment. NM: guarantor, writing, draft editing, public involvement design, tool design.

Funding Funder: Medical Research Council; Award: MICA: A partnership to extend the research utility of a source of real-world health data, the UK National Neonatal Research Database; Reference: MR/T016752/1; Chief Investigator: NM.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by The National Neonatal Research Database (NNRD) and by the UK National Research Ethics Service as a database (10/80803/151) and the Health Research Authority Confidentiality Advisory Group (8-05(f)/0210). Research ethics approval for PPI consultations is not required, per NIHR guidelines. However, we approached parents and former patients through the neoWONDER group who had agreed to be invited to participate in consultations (REC reference: 20/yh/0330). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iD

William Bishop Lammons http://orcid.org/0000-0002-8302-9831

REFERENCES

1 Gale C, Morris I, Neonatal Data Analysis Unit (NDAU) Steering Board. The UK National Neonatal Research Database: using neonatal data for research, quality improvement and more. *Arch Dis Child Educ Pract Ed* 2016;101:216–8.
2 Wong HS, Hopkins L, O'Donovan MC, *et al*. Pilot study to establish a prospective neonatal cohort: study of preterm infants and neurodevelopmental genes (SPRING). *BMJ Paediatr Open* 2020;4:e000648.
3 Neonatal Data Analysis Unit (NDAU). Neonatal Health Intelligence Tool. 2021. Available: https://www.imperial.ac.uk/neonatal-data-analysis-unit/neonatal-data-analysis-unit/neonatal-data-visualisations/
4 INVOLVE, NIHR. Patient and public involvement in research and research ethics committee review. NIHR version 1. 2009. Available: https://www.invo.org.uk/wp-content/uploads/2011/12/INVOLVENRESfinalStatement310309.pdf
5 INVOLVE, NIHR (National Institute of Health Research). Public involvement in research: impact on ethical aspects of research. 2014. Available: www.involve.nihr.ac.uk
6 National Institute for Health Research. Patient and public involvement in health and social care research: a handbook for researchers. London Research Design Service; 2014. 1–40.Available: http://www.rds.nihr.ac.uk/wp-content/uploads/RDS-PPI-Handbook-2014-v8-FINAL.pdf
7 Morris RL, Ruddock A, Gallacher K, *et al*. Developing a patient safety guide for primary care: a co-design approach involving patients, carers and clinicians. *Health Expect* 2021;24:42–52.
8 Bentley S, Johnson C, Exall E, *et al*. Improving patient-clinician communication following nephrectomy in renal cell carcinoma: development, content validation and pilot testing of a conversation aid tool. *Patient Educ Couns* 2021;104:S0738-3991(20)30347-5:99–108.:.
9 Lammons W, Moss B, Battersby C, *et al*. Incorporating parent, former patient and clinician perspectives in the design of a national UK double-cluster, randomised controlled trial addressing uncertainties in preterm nutrition. *BMJ Paediatr Open* 2021;5:e001112.
10 Husbands S, Jowett S, Barton P, *et al*. Understanding and identifying key issues with the involvement of clinicians in the development of decision-analytic model structures: a qualitative study. *Pharmacoeconomics* 2018;36:1453–62.
11 Tang T, Lim ME, Mansfield E, *et al*. Clinician user involvement in the real world: designing an electronic tool to improve interprofessional communication and collaboration in a hospital setting. *Int J Med Inform* 2018;110:S1386-5056(17)30425-2:90–7.:.
12 UK Standards for Public Involvement, NIHR. UK public involvement standards development partnership. 2019. Available: https://sites.google.com/nihr.ac.uk/pi-standards/home
13 neoWONDER. neoWONDER: neonatal whole population data linkage to improve lifelong health and wellbeing of preterm babies 2020. 2020. Available: https://www. Neowonder.org.uk/
14 Muller I, Santer M, Morrison L, *et al*. Combining qualitative research with ppi: reflections on using the person-based approach for developing behavioural interventions. *Res Involv Engagem* 2019;5:34.
15 Scott IA, Carter SM, Coiera E. Exploring stakeholder attitudes towards AI in clinical practice. *BMJ Health Care Inform* 2021;28:e100450.
16 Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods* 2006;5:1–11.
17 Ritchie J. *Qualitative research practice: a guide for social science students and researchers*. Thousand Oaks, CA: Sage Publications, 2003: 24–6.
18 Inductive and deductive coding and theme development. *Int J Qual Methods* 2006;5:80–92.
19 Ritchie J, Lewis J, Elam G. Designing and selecting samples. In: Ritchie J, Lewis J, eds. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. Thousand Oaks, CA: Sage Publications Sage CA, n.d.: 77–108.
20 Ritchie J, Spencer L, O'Connor W. Carrying out qualitative analysis. In: Ritchie J, Lewis J, eds. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. Thousand Oaks, CA: Sage Publications, n.d.: 219–62.
21 Spencer L, Ritchie J, O'Connor William. Analysis: practices, principles and processes. In: Ritchie J, Lewis J, eds. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. Thousand Oaks, CA: Sage Publications, n.d.: 199–218.
22 Staniszewska S, Brett J, Simera I, *et al*. GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. *Res Involv Engagem* 2017;3:13.
23 Kueper JK, Terry A, Bahniwal R, *et al*. Connecting artificial intelligence and primary care challenges: findings from a multi stakeholder collaborative consultation. *BMJ Health Care Inform* 2022;29:e100493. 10.1136/bmjhci-2021-100493 Available: https://doi.org/10.1136/bmjhci-2021-100493
24 Ryan BL, Brown JB, Terry A, *et al*. Implementing and using a patient portal: a qualitative exploration of patient and provider perspectives on engaging patients. *J Innov Health Inform* 2016;23:848. 10.14236/jhi.v23i2.848 Available: http://dx.doi.org/10.14236/jhi.v23i2.848
25 Cresswell K, Callaghan M, Mozaffar H, *et al*. NHS Scotland's decision support platform: a formative qualitative evaluation. *BMJ Health Care Inform* 2019;26:e100022:1–9.:.
26 Austin EJ, Lee JR, Ko CW, *et al*. Improving the impact of clinical documentation through patient-driven co-design: experiences with cancer pathology reports. *BMJ Health Care Inform* 2020;27:e100197.

BMJ Health & Care Informatics

Accuracy of a tool to prioritise patients awaiting elective surgery: an implementation report

Videha Sharma ,[1] Rowan Pritchard-Jones,[2] Sharon Scott,[3] John Ainsworth[1]

To cite: Sharma V, Pritchard-Jones R, Scott S, *et al*. Accuracy of a tool to prioritise patients awaiting elective surgery: an implementation report. *BMJ Health Care Inform* 2023;**30**:e100687. doi:10.1136/bmjhci-2022-100687

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjhci-2022-100687).

Received 05 October 2022
Accepted 06 January 2023

[1]Centre for Health Informatics, Division of Informatics, Imaging and Data Science, University of Manchester Faculty of Biology, Medicine and Health, Manchester, UK
[2]St Helens and Knowsley Teaching Hospitals NHS Trust, Prescot, UK
[3]Liverpool University Hospitals NHS Foundation Trust, Liverpool, UK

Correspondence to
Mr Videha Sharma;
videha.sharma@manchester.ac.uk

ABSTRACT

Study objective The objective of this study was to evaluate the accuracy of a new elective surgery clinical decision support system, the 'Patient Tacking List' (PTL) tool (C2-Ai(c)) through receiver operating characteristic (ROC) analysis.

Methods We constructed ROC curves based on risk predictions produced by the tool and compared these with actual patient outcomes on a retrospective cohort of patients awaiting elective surgery.

Results A total of 11 837 patients were included across three National Health Service (NHS) hospitals in England. ROC analysis revealed an area under the curve of 0.95 (95% CI 0.92 to 0.98) for mortality and 0.8 (95% CI 0.78 to 0.82) for complications.

Discussion The PTL tool was successfully integrated into existing data infrastructures, allowing real-time clinical decision support and a low barrier to implementation. ROC analysis demonstrated a high level of accuracy to predict the risk of mortality and complications after elective surgery. As such, it may be a valuable adjunct in prioritising patients on surgical waiting lists.
Health systems, such as the NHS in England, must look at innovative methods to prioritise patients awaiting surgery in order to best use limited resources. Clinical decision support tools, such as the PTL tool, can improve prioritisation and thus positively impact clinical care and patient outcomes.

Conclusions The high level of accuracy for predicating mortality and complications after elective surgery using the PTL tool indicates the potential for clinical decision support tools to help tackle rising waiting lists and improve surgical planning.

INTRODUCTION

Elective waiting lists for surgery in England are currently stratified based on a prioritisation system produced by the Academy of Royal Colleges and endorsed by the National Health Service (NHS). The system relies on healthcare professionals, typically a surgeon, to manually assign a priority code (P-code) to each patient listed for elective surgery within their domain (P1 (highest) - P4 (lowest)). Following assignment of a P-code, the patient is expected to undergo surgery within a stipulated time frame, for example, the assigned code P3 means the patient must undergo surgery within 3 months. Stepping outside of these time frames means patients should be subjected to a harm review.[1]

The COVID-19 pandemic has widely disrupted the delivery of healthcare services, including elective surgery.[2 3] As a result, the number of patients awaiting surgery has sharply risen, which is sometimes referred to as the 'elective backlog'.[4] The current method to prioritise patients is procedure-specific and simplified to allow rapid prioritisation. It is not designed to manage the priority within a group of P-coded patients. Yet, there will be those who deteriorate faster than others due to their pattern of comorbidities. There is a need to improve the accuracy of assessing patients listed for elective surgery and prioritise based on greater objectivity. As such, digital tools to improve this process have been proposed, such as the use of predictive algorithms and artificial intelligence. These could have a positive impact on identifying patients at greatest risk of harm from waiting for a procedure and thus improve clinical care and outcomes. With the electronic management of elective waiting lists, this potential may now be realised; however, active intelligent management with clinical decision support tools has not been reported outside of research settings.[5 6]

In the field of predicting the risk of mortality and complications in surgery, the Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity (POSSUM) scoring system is one of the most established and widely accepted constructs.[7] It has been iterated over time and been shown to be highly accurate at predicting adverse outcomes and death in a range of surgical procedures across specialities.[8] The variables used to power POSSUM include routinely

collected demographic and clinical parameters, which can be found in online supplemental file 1.

The Patient Tacking List (PTL) tool (C2-Ai(c)) is a clinical decision support tool based on the POSSUM Score that can help prioritise patients on an elective waiting list. The tool combines POSSUM, the planned surgical procedure details and time on the waiting list, and applies these to a referential dataset to produce a 'matrix score'. The matrix score represents the difference in mortality and complication rate between the procedure being done electively versus waiting for the patient to decompensate and present as an emergency. Matrix scores range between 4 and 100, with a higher score corresponding to a greater risk of a poor outcome if the procedure is not done electively. Users of the tool can visualise their waiting list on a bespoke user-interface, which orders patients based on matrix scores, with options to filter based on demographics, clinical specialty and procedure type. These visualisations give a risk-stratified overview of patients awaiting surgery allowing better informed surgical planning.

The PTL tool has the potential to improve the accuracy of elective surgical risk-stratification and support clinical prioritisation. Furthermore, by including objective measures of risk, it may reduce variation, and thus improve equity in patient care. However, no objective measure of the accuracy of the tool to predict mortality and complications has previously been undertaken.

This report describes the implementation of the PTL tool as a pilot solution at three NHS trusts. The tool was used to analyse patients listed for elective surgery and produce matrix scores alongside current standard practice.

AIMS

The aim of this study is to evaluate the accuracy of the PTL tool through receiver operating characteristic (ROC) analysis for mortality and complications in a retrospective cohort.

METHODS

The tool was deployed in March 2021 and this study used data for the subsequent 12 months across three NHS Foundation Trusts including St Helens and Knowsley, Warrington and Halton and Royal Liverpool and Aintree Hospitals. The tool has subsequently been integrated within the regional clinical data warehouse (Combined Intelligence for Population Health Action) with the intention to process data and produce matrix scores for all patients listed for elective surgery. Data on patient comorbidities were accessed using a download of 2 years complete Secondary Use Service data (NHS digital) in all specialties.

We constructed ROC curves for mortality and complications based on PTL risk predictions and compared these with actual patient outcomes to assess the accuracy of the tool. ROC curves are a graphical approach to evaluate the connection/trade-off between clinical sensitivity and specificity for every possible cut-off for a test or a combination of tests. The area under the ROC curve describes the potential benefit of using the test(s) in question.

RESULTS

A total of 11 837 patients were included in the retrospective analysis. The outcomes for patients undergoing surgery between March 2021 and March 2022 using the predictions from the PTL tool showed an area under the ROC curve of 0.95 (95% CI 0.92 to 0.98) for mortality and 0.8 (95% CI 0.78 to 0.82) for overall complications (figure 1). Anecdotally, a 15 min saving of surgeon time was reported per patient each time the waiting list was reprioritised (P-coded).

DISCUSSION

Principle findings

This study found that the PTL tool accurately predicted the risk of mortality and complications for patients listed for elective surgery. Matrix scores correlated well to potential adverse outcomes and may therefore be used to prioritise patients on surgical waiting lists.

Implications for clinical practice

Planning elective surgical waiting lists and prioritising patients to optimise the utilisation of resources is a complex undertaking. Several tools to support this have been reported in the literature. A recent systematic review by Dery *et al*[9] identified 34 different tools. Most included studies reported on the development or clinical validation of the tool—rather than the implementation into clinical practice. The authors of this review concluded that implementation into clinical practice remains a challenge.[9] Our implementer report describes the real-world application of a prioritisation tool in the NHS. It has demonstrated the ability to identify patients at greater risk of adverse outcomes and therefore expedite their waiting time for surgery. The limited integration required lowered the barrier to implementation and demonstrated the potential scalability of this tool.

Previous studies have shown that by risk stratifying patients, there is an opportunity to personalise management, optimise prehabilitation and improve postoperative outcomes, such as a reduction in pulmonary complications.[10] Patients with higher matrix scores identified through the PTL tool may thus be selected for prehabilitation, with a view of reducing their risk while they wait for surgery. Future studies may explore the impact on postoperative outcomes of patients with higher matrix who underwent prehabilitation to provide a measure of the impact of the tool.

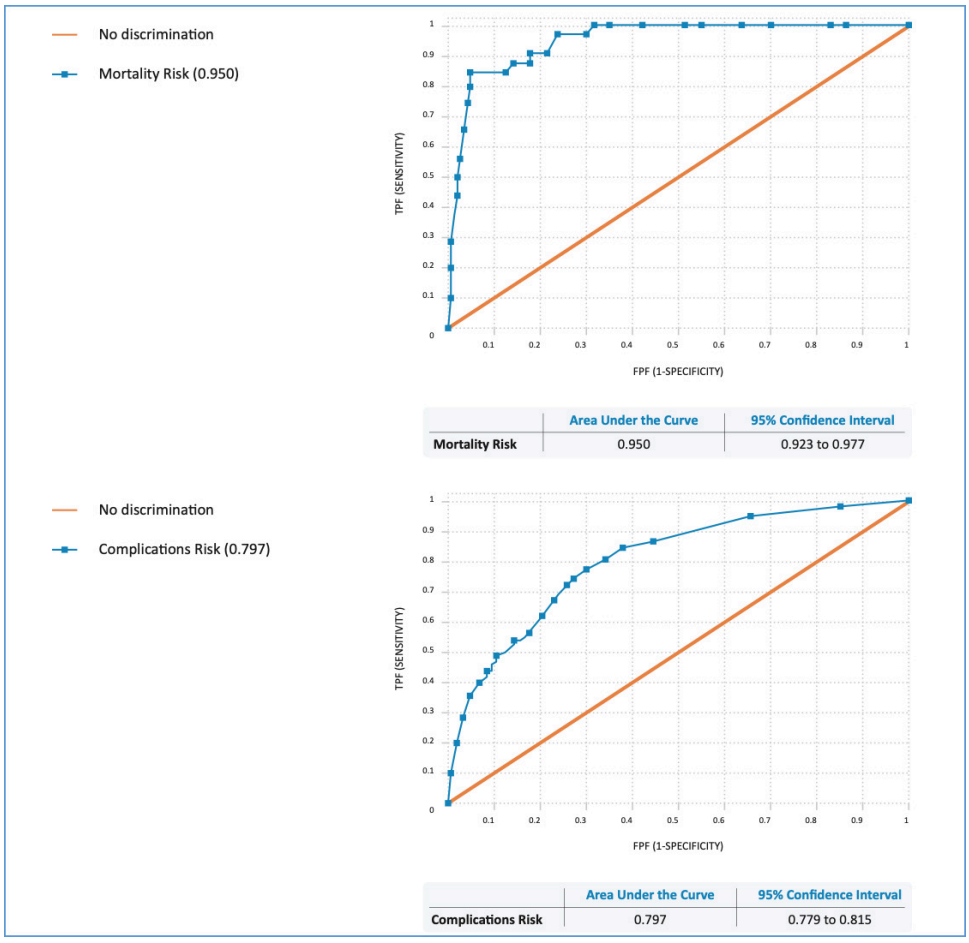

Figure 1 Receiver operating characteristic curves for mortality and overall complications. FPF, false positive fraction; TPF, true positive fraction.

Given the increasing complexity in healthcare and larger number of patients on waiting lists, the tool has the potential to lighten the administrative burden and reduce costs related to service planning and delivery. To achieve maximum impact from such implementations, the importance of utility must be shared with all stakeholders including healthcare professionals, administrative staff and patients. This will result in the wider cultural change associated with digital transformation. Further work to evaluate the usability of the tool is thus warranted, to better understand users' experience, integration into administrative workflows and identify areas for improvement.

CONCLUSION

Clinical decision support systems, such as the PTL tool, can improve the prioritisation of patients requiring elective surgery. This can improve overall mortality and complications related to surgical conditions and positively impact elective backlog by accurately allocating healthcare resources.

Twitter Videha Sharma @VidehaSharma

Contributors VS and RP-J conceptualised the study. VS undertook statistical analysis and wrote the first draft of the manuscript. RP-J, SS and JA reviewed and edited the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests RP-J has previously been supported by C2Ai to attend an innovation conference and a digital health awards event for work related to the tool in this report. All other authors declare that they have no conflicts of interest.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

ORCID iD
Videha Sharma http://orcid.org/0000-0001-7640-1239

REFERENCES

1 Oudhoff JP, Timmermans DRM, Knol DL, *et al*. Prioritising patients on surgical waiting lists: a conjoint analysis study on the priority judgements of patients, surgeons, occupational physicians, and general practitioners. *Soc Sci Med* 2007;64:1863–75.
2 Peek N, Sujan M, Scott P. Digital health and care in pandemic times: impact of COVID-19. *BMJ Health Care Inform* 2020;27:e100166.
3 COVIDSurg Collaborative. Elective surgery cancellations due to the COVID-19 pandemic: global predictive modelling to inform surgical recovery plans. *Br J Surg* 2020;107:1440–9.
4 Macdonald N, Clements C, Sobti A, *et al*. Tackling the elective case backlog generated by Covid-19: the scale of the problem and solutions. *J Public Health* 2020;42:712–6.
5 Everett JE. A decision support simulation model for the management of an elective surgery waiting system. *Health Care Manag Sci* 2002;5:89–95.
6 Dios M, Molina-Pariente JM, Fernandez-Viagas V, *et al*. A decision support system for operating room scheduling. *Comput Ind Eng* 2015;88:430–43.
7 Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. *Br J Surg* 1991;78:355–60.
8 Copeland G. Assessing the surgeon: 10 years experience with the POSSUM system. *Journal of Clinical Excellence* 2000;2:187–90.
9 Déry J, Ruiz A, Routhier F, *et al*. A systematic review of patient prioritization tools in non-emergency healthcare services. *Syst Rev* 2020;9:1–14.
10 Hughes MJ, Hackney RJ, Lamb PJ, *et al*. Prehabilitation before major abdominal surgery: a systematic review and meta-analysis. *World J Surg* 2019;43:1661–8.

BMJ Health & Care Informatics

Evaluation of race/ethnicity-specific survival machine learning models for Hispanic and Black patients with breast cancer

Jung In Park ,[1] Selen Bozkurt,[2] Jong Won Park,[3] Sunmin Lee[4]

To cite: Park JI, Bozkurt S, Park JW, *et al*. Evaluation of race/ethnicity-specific survival machine learning models for Hispanic and Black patients with breast cancer. *BMJ Health Care Inform* 2023;**30**:e100666. doi:10.1136/bmjhci-2022-100666

Received 18 August 2022
Accepted 29 December 2022

[1]Sue & Bill Gross School of Nursing, University of California Irvine, Irvine, California, USA
[2]Stanford University, Stanford, California, USA
[3]Yonsei University College of Medicine, Seoul, Seodaemun-gu, Korea (the Republic of)
[4]School of Medicine, University of California Irvine, Irvine, California, USA

Correspondence to
Dr Jung In Park;
junginp@uci.edu

ABSTRACT

Objectives Survival machine learning (ML) has been suggested as a useful approach for forecasting future events, but a growing concern exists that ML models have the potential to cause racial disparities through the data used to train them. This study aims to develop race/ethnicity-specific survival ML models for Hispanic and black women diagnosed with breast cancer to examine whether race/ethnicity-specific ML models outperform the general models trained with all races/ethnicity data.

Methods We used the data from the US National Cancer Institute's Surveillance, Epidemiology and End Results programme registries. We developed the Hispanic-specific and black-specific models and compared them with the general model using the Cox proportional-hazards model, Gradient Boost Tree, survival tree and survival support vector machine.

Results A total of 322 348 female patients who had breast cancer diagnoses between 1 January 2000 and 31 December 2017 were identified. The race/ethnicity-specific models for Hispanic and black women consistently outperformed the general model when predicting the outcomes of specific race/ethnicity.

Discussion Accurately predicting the survival outcome of a patient is critical in determining treatment options and providing appropriate cancer care. The high-performing models developed in this study can contribute to providing individualised oncology care and improving the survival outcome of black and Hispanic women.

Conclusion Predicting the individualised survival outcome of breast cancer can provide the evidence necessary for determining treatment options and high-quality, patient-centred cancer care delivery for under-represented populations. Also, the race/ethnicity-specific ML models can mitigate representation bias and contribute to addressing health disparities.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Survival machine learning allows healthcare professionals to identify patients at high risk, but models trained with data poorly representative of minority groups, they may exacerbate health disparities. To date, no study developed race/ethnicity-specific survival machine learning models for Hispanic and black women diagnosed with breast cancer.

WHAT THIS STUDY ADDS

⇒ The race/ethnicity-specific survival machine learning models outperformed the general models trained with all races/ethnicity when predicting the outcomes of specific races/ethnicity.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Predicting the individualised survival outcome of breast cancer can provide the evidence necessary for determining treatment options and high-quality, patient-centred cancer care delivery for underrepresented populations. Also, the race/ethnicity-specific machine learning models can mitigate representation bias and contribute to addressing health disparities.

INTRODUCTION

Breast cancer is the second-leading cause of cancer-related deaths in women in the USA, and it affects every ethnic group of women in the USA.[1 2] However, there are racial and ethnic divides in cancer survival. Breast cancer is the most prevalent reason for cancer-related death in Hispanic women in the USA.[3] Also, minority women, especially black women, have a higher mortality rate (26.8 per 100 000 women) even though white women (18.8 per 100 000 women) have higher cancer incidence.[2 4 5] These facts indicate that the cancer survival rates need to be improved among Hispanic and black women, and various features contributing to breast cancer mortality should be understood to provide tailored intervention for enhanced survival.

Unlike traditional survival models that use a standard statistical method, survival machine learning (ML) has been suggested as a useful approach for learning the patterns from high-dimensional data and complex feature interactions for forecasting future events.[6] This approach allows healthcare professionals to identify patients at high risk or predict those

who need increased utilisation of healthcare services to proactively support and provide interventions necessary for the patients.[7] However, a growing concern exists that ML models have the potential to cause racial disparities through the data used to train them.[8] The ML model trained with the data representing general population would not contain sufficient number of participants from the minority population and is biased, resulting in inaccurate predictions for the minority group even if the overall accuracy is high.[9] If the ML models trained with data poorly representative of minority groups are used in healthcare, they may exacerbate health disparities.[10] To address such harmful effects, it is recommended to train an ML model with data that resemble the population that the model is intended to use.[11 12] To the best of our knowledge, no study developed race/ethnicity-specific survival ML models for Hispanic and black women diagnosed with breast cancer.

Therefore, there is a need for race/ethnicity-specific survival ML models trained with the underrepresented populations to examine the feasibility of race/ethnicity-specific ML models that may outperform the general model trained with all races/ethnicity. Accurate prediction of the individualised outcome will enable tailored healthcare delivery and a better outcome for the underrepresented populations. This study aims to develop race/ethnicity-specific survival ML models for Hispanic and black women diagnosed with breast cancer to examine whether race/ethnicity-specific ML models outperform the models trained with the general population data when predicting the survival of Hispanic and black women diagnosed with breast cancer.

METHODS

Data source

We used the data from the US National Cancer Institute's population-based Surveillance, Epidemiology and End Results (SEER) programme registries. The SEER programme currently collects and publishes cancer incidence and survival data in the USA from population-based cancer registries in 22 geographical areas, representing approximately 48% of the US population.[13] The SEER data are considered the gold standard for data quality among cancer registries in the USA and globally.[14] We selected adult female patients' data (18 or older) from SEER who had breast cancer diagnoses between 1 January 2000 and 31 December 2017. Also, we selected California as the geographical location for the diverse characteristics of the patient population. The Hispanic population included all races, and the black population was non-Hispanic. Figure 1 shows the flow chart of data collection.

Predictor and outcome variables

The predictor variables included age at cancer diagnosis, marital status at diagnosis, first malignant primary tumour indicator, the sequence number of tumours, primary site, histology, the total number of in situ/malignant tumours,

Figure 1 Flow chart of data collection.

SEER summary stage, derived stage, grade, regional lymph nodes examined, regional lymph nodes positive, oestrogen receptor status, progesterone receptor status, chemotherapy, radiation, sequence of radiation and surgery performed, reason no cancer-directed surgery and sequence of systemic therapy and surgical procedures. Vital status was recorded as alive/dead at the time of the cut-off date (31 December 2017). The sequence number of tumours describes the sequence of all reportable tumours that occurred over a patient's lifetime.

The outcome variable was the survival months of a patient.

Data preprocessing and preparation

Before training the survival models, we preprocessed the predictor variables to enhance the ML modelling performance. Rows containing missing values were dropped. All the categorical features were reencoded using a one-hot-encoding scheme where each new column represented a single category. We applied variance filtering (with the threshold of 0.01) to drop the features that were near-constant or had low variance. Thus, a feature containing outliers would appear as a low-variance column and be filtered out. Once the preprocessing was completed, the final dataset was exported into a new flat file for the training. To train an ML model for survival analysis, the 'survival months' variable was used as the target for the training. 'Vital status' was used for the event.

We took several steps for data preparation to develop race/ethnicity-specific models for the Hispanic and black populations and compare them with the general model that included all races/ethnicity. Figure 2 shows the process of data preparation for model development.

First, we split the full dataset into a training set (T_{all}) for model development and a test set (E_{all}) for evaluation with a 7:3 ratio to randomly sample the populations. Each set was used to sample the populations for model

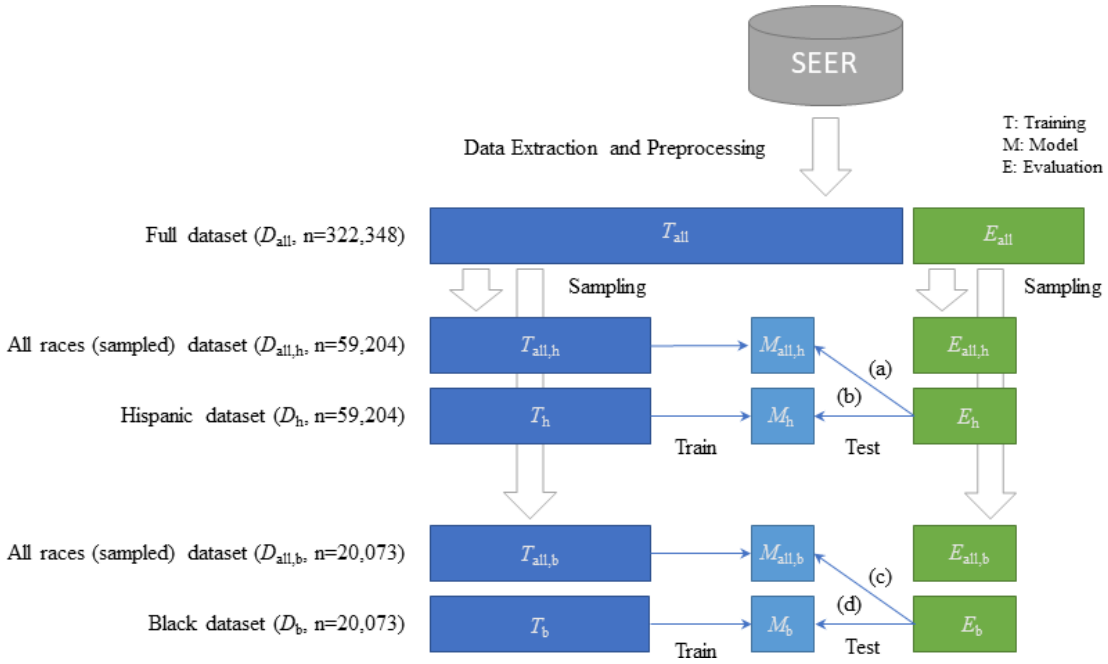

Figure 2 Data preparation for model development. SEER, Surveillance, Epidemiology and End Results.

development randomly. The randomly sampled population sets maintained the original ratio of each race/ethnicity in the full dataset. Second, we extracted the Hispanic population from the original training set, T_{all} (T_h) to train the Hispanic-specific model (M_h). We also extracted the Hispanic population from the original test set, E_{all} (E_h) to test the model, M_h. Then, we randomly sampled the populations from the original training set (T_{all}) that included all races/ethnicity ($T_{all,h}$), to match the exact number of samples used for the Hispanic-specific model training. We also randomly sampled the populations from the original test set (E_{all}) that included all races/ethnicity ($E_{all,h}$), to match the exact number of samples used for the Hispanic-specific model testing. $T_{all,h}$ was used to develop a model $M_{all,h}$. Then, the performance of the models $M_{all,h}$ (a) and M_h (b) were compared with the same test set, E_h. Third, we repeated the process of Hispanic-specific model development for Black-specific model development.

We extracted the black population from the original training set, T_{all} (T_b) to train the black-specific model (M_b). We also extracted the black population from the original test set, E_{all} (E_b) to test the model, M_b. Then, we randomly sampled the populations from the original training set (T_{all}) that included all races/ethnicity ($T_{all,b}$), to match the exact number of samples used for the black-specific model training. We also randomly sampled the populations from the original test set (E_{all}) that included all races/ethnicity ($E_{all,b}$), to match the exact number of samples used for the black-specific model testing. $T_{all,b}$ was used to develop a model $M_{all,b}$. Then, the performance of the models $M_{all,b}$ (c) and M_b (d) were compared with the same test set, E_b.

Race/ethnicity-specific models

For the survival ML modelling, we developed and compared four models: Cox proportional-hazards (PH) model (CoxPH), Gradient Boost Tree (GBT), survival

Table 1 Description of survival machine learning models

Model	Description
Cox PH	A standard survival model looking at the effects of a patient's covariates on the risk of death.[26] It is a multivariate regression model for survival analysis.[27]
GBT	An ensemble learning method that sequentially combines the outputs from individual decision trees, so each new tree can predict and correct the errors of the previous tree.[28] It uses Gradient-boosted Cox proportional hazard loss with regression trees as base learner.[29]
ST	A model that splits the covariate space into smaller nodes containing observations with homogeneous survival outcomes.[30] It is a tree-based method for censored survival data.[31]
SSVM	An extension of the standard SVM to maximise the concordance index (*C-index*) and account for complex, non-linear relationships between features and survival.[32] It is an efficient way of training a kernel SVM.[33]

GBT, Gradient Boost Tree; PH, proportional hazard; SSVM, survival support vector machine; ST, survival tree.

tree (ST) and survival support cector machine (SSVM). The description of each model is shown in table 1.

Each model's performance was evaluated using the C-index. The C-index is a standard way of measuring the performance of survival models. It can be viewed as the fraction of all pairs of patients predicted to have correct orders over the total number of possible evaluation pairs.[15]

For each race/ethnicity, we trained and compared two different models based on the two datasets mentioned above—one with a specific race/ethnicity and the other one with all races/ethnicity. Our hypothesis was that the model trained with specific race/ethnicity would outperform the general model trained with all races/ethnicity when predicting the breast cancer survival of a specific race/ethnicity.

RESULTS

Sample characteristics

A total of 322 348 female patients who had breast cancer diagnoses between 1 January 2000 and 31 December 2017 were identified. Among them, the number of Hispanic patients was 59 204 (18.4%), and black was 20 073 (6.2%). Table 2 shows the detailed characteristics of the study sample, Hispanic, black and all races/ethnicity.

Compared with all races/ethnicity (15.2%) and Hispanic (14.9%) populations, more black population was dead (24.4%). Hispanic population's survival months (mean: 80.6, median: 67.0) were lower compared with all races/ethnicity (90.4, 79.0) and black (82.9, 69.0) populations. Hispanic population was younger (mean:55.3, median 54.0), compared with all races/ethnicity (59.1, 59.0) and black (57.8, 57.0) populations. Black (36.6%) and Hispanic (39.2%) population had higher percentage of poorly differentiated grade III cancer, compared with all races/ethnic (32.7%) groups. Black population had lower percentages of positive oestrogen receptor status (65.6%), compared with all races/ethnicity (77.4%) and Hispanic (73.5%) populations. Also, black population had lower percentages of positive progesterone receptor status (52.1%), compared with all races/ethnicity (65.6%) and Hispanic (62.3%) populations.

Lower percentages of Hispanic (51.0%) and black (50.0%) populations had chemotherapy compared with all races/ethnicity (57.4%). Higher percentages of black (57.3%) and Hispanic (55.6%) populations had no radiation and/or cancer-directed surgery, compared with all races/ethnicity (52.3%). Higher percentages of overall (46.4%) and Hispanic (43.4%) populations had radiation after surgery than Black (41.5%) populations.

Figure 3 shows the Kaplan-Meier curves of the Hispanic, black and all races/ethnic groups. All races/ethnic groups had the better survival than the Hispanic and black groups.

Data preprocessing and preparation

After data preprocessing and cleaning, the final dataset for analysis contained 260 variables. Values in 'Derived stages' variable were grouped into '0', 'I', 'II', 'III', 'IV' and 'unknown'. 'Regional nodes examined' and 'Regional nodes positive' were integer variables which contains both numeric and encoded values (90+). Numeric values were categorised (ie, 0–9, 11–19, …, 40+), while encoded values were mapped to 'oher'.

Model development

We extracted 59 204 Hispanic populations for each training set for Hispanic-specific model (T_h) and a comparison model with all races/ethnicity ($T_{all,h}$). Also, we extracted 20 073 black populations for each training set for black-specific model (T_b) and a comparison model with all races/ethnicity ($T_{all,b}$). Once data were prepared, we applied variance filtering and dropped the features that had low variance. After filtering, the number of features we had for the T_h was 72, and for the $T_{all,h}$ was 71 for Hispanic-specific model training, and the number of features we had for the T_b and $T_{all,b}$ was 72 for the black-specific model training.

During the training, both training sets (race-specific and all races/ethnicity) were further split into actual training set and validation set during a cross-validation phase when parameter tuning was necessary (GBT and ST models). We used random search method to find the most optimal parameters for each survival analysis model. We used 20 iterations and 5-fold cross validation was used for all cases for each training. We used scikit-survival package (V.0.17.1) for the modelling (CoxPHSurvivalAnalysis class for CoxPH, Gradient Boosting Survival Analysis class for GBT, SurvivalTree class for ST and FastKernelSurvivalSVM class for SSVM), scikit-learn (V.1.0.2) for the feature selection (VarianceThreshold), hyperopt (V.0.2.7) for the hyperparameter search, and pandas (V.1.4.1) for general data preprocessing and preparation.

Model evaluations

The model evaluation results are shown in table 3 and figure 4 where we compared different combinations of modelling methods and input training/test sets.

Hispanic-specific model (M_h) and all races/ethnicity model ($M_{all,h}$) were evaluated using the same the test set (E_h). Hispanic-specific model (M_h) outperformed all races/ethnicity model ($M_{all,\ h}$) in three out of four approaches, which were Cox PH (0.832 vs 0.828), ST (0.772 vs 0.763) and SSVM (0.834 vs 0.790). The GBT model showed the same c-index score (0.813) for both models.

Black-specific model (M_b) and all races/ethnicity model ($M_{all,b}$) were evaluated using the same the test set (E_b). Black-specific model (M_b) outperformed all races/ethnicity model ($M_{all,b}$) in all four approaches, Cox PH (0.823 vs 0.821), GBT (0.808 vs 0.803), ST (0.804 vs 0.801) and SSVM (0.824 vs 0.786). In both race/

Table 2 Sample characteristics

Non-Hispanic (NH) white	199 913 (62.0)		
Hispanic (all races)	**59 204 (18.4)**		
NH Asian or Pacific Islander	41 811 (13.0)		
NH black	**20 073 (6.2)**		
NH American Indian/Alaska Native	1347 (0.4)		
Total	322 348		
	All races/ethnicity (N=322 348)	**Hispanic (N=59 204)**	**Black (N=20 073)**
Vital status (n, %)			
Alive	273 455 (84.8)	50 382 (85.1)	15 175 (75.6)
Dead	48 893 (15.2)	8822 (14.9)	4898 (24.4)
Survival months			
Min, Max	0.0, 227.0	0.0, 227.0	0.0, 227.0
Mean	90.4	80.6	82.9
Median	79.0	67.0	69.0
SD	60.9	58.7	59.8
Age			
Min, Max	19.0, 100.0	19.0, 99.0	19.0, 100.0
Mean	59.1	55.3	57.8
Median	59.0	54.0	57.0
SD	13.0	12.9	13.0
First malignant primary indicator (n, %)			
Y	278 117 (86.3)	53 174 (89.8)	17 240 (85.9)
N	44 231 (13.7)	6030 (10.2)	2833 (14.1)
Sequence no of tumours (n, %)			
One primary only	236 182 (73.3)	46 956 (79.3)	14 650 (73.0)
Second of two or more primaries	44 226 (13.7)	6233 (10.5)	2812 (14.0)
First of two or more primaries	34 893 (10.8)	5328 (9.0)	2152 (10.7)
Third of three or more primaries	6008 (1.9)	603 (1.0)	397 (2.0)
Fourth of four or more primaries	878 (0.3)	75 (0.1)	51 (0.3)
Other	161 (0.0)	9 (0.0)	11 (0.1)
Histology (n, %)			
Ductal and lobular neoplasms	306 634 (95.1)	56 417 (95.3)	18 904 (94.2)
Cystic, mucinous and serous neoplasms	5800 (1.8)	1001 (1.7)	411 (2.0)
Adenomas and adenocarcinomas	5063 (1.6)	768 (1.3)	294 (1.5)
Epithelial neoplasms, NOS	1343 (0.4)	279 (0.5)	139 (0.7)
Complex epithelial neoplasms	1311 (0.4)	258 (0.4)	154 (0.8)
Adnexal and skin appendage neoplasms	807 (0.3)	136 (0.2)	65 (0.3)
Squamous cell neoplasms	573 (0.2)	116 (0.2)	52 (0.3)
Fibroepithelial neoplasms	385 (0.1)	141 (0.2)	19 (0.1)
Other	432 (0.1)	88 (0.1)	35 (0.2)
Total no of in situ/malignant tumours (n, %)			
1	239 762 (74.4)	47 536 (80.3)	14 880 (74.1)
2	67 146 (20.8)	10 046 (17.0)	4224 (21.0)
3	12 622 (3.9)	1387 (2.3)	802 (4.0)
4	2276 (0.7)	206 (0.3)	141 (0.7)

Continued

Table 2 Continued

5+	542 (0.2)	29 (0.0)	26 (0.1)
Summary stage (n, %)			
Localised	205 612 (63.8)	34 061 (57.5)	11 398 (56.8)
Regional	102 914 (31.9)	22 292 (37.7)	7291 (36.3)
Distant	13 822 (4.3)	2851 (4.8)	1384 (6.9)
Grade (n, %)			
Moderately differentiated; grade II	138 937 (43.1)	24 491 (41.4)	9346 (46.6)
Poorly differentiated; grade III	105 361 (32.7)	23 237 (39.2)	7350 (36.6)
Well differentiated; grade I	73 815 (22.9)	10 607 (17.9)	3012 (15.0)
Undifferentiated; anaplastic; grade IV	4235 (1.3)	869 (1.5)	365 (1.8)
Regional lymph nodes examined (n, %)			
0–9	233 642 (72.5)	39 268 (66.3)	13 679 (68.1)
10–19	61 685 (19.1)	13 312 (22.5)	4470 (22.3)
20–29	17 330 (5.4)	4167 (7.0)	1163 (5.8)
Other	6275 (1.9)	1547 (2.6)	549 (2.7)
30–39	2837 (0.9)	740 (1.2)	176 (0.9)
40+	579 (0.2)	170 (0.3)	36 (0.2)
Oestrogen receptor status (n, %)			
Positive	249 656 (77.4)	43 494 (73.5)	13 168 (65.6)
Negative	56 876 (17.6)	12 575 (21.2)	5941 (29.6)
Borderline/unknown	15 071 (4.7)	2925 (4.9)	915 (4.6)
N/A	745 (0.2)	210 (0.4)	49 (0.2)
Progesterone receptor status (n, %)			
Positive	210 985 (65.5)	36 867 (62.3)	10 463 (52.1)
Negative	90 699 (28.1)	18 369 (31.0)	8224 (41.0)
Borderline/unknown	19 919 (6.2)	3758 (6.3)	1337 (6.7)
N/A	745 (0.2)	210 (0.4)	49 (0.2)
Chemotherapy (n, %)			
Y	185 019 (57.4)	30 197 (51.0)	10 038 (50.0)
N	137 329 (42.6)	29 007 (49.0)	10 035 (50.0)
Radiation (n, %)			
None/unknown	150 350 (46.6)	29 136 (49.2)	9965 (49.6)
Beam radiation	147 114 (45.6)	25 651 (43.3)	8429 (42.0)
Recommended, unknown if administered	10 290 (3.2)	2532 (4.3)	922 (4.6)
Refused	5901 (1.8)	796 (1.3)	403 (2.0)
Radioactive implants	5897 (1.8)	645 (1.1)	191 (1.0)
Radiation, NOS method or source not specified	2415 (0.7)	367 (0.6)	150 (0.7)
Other	381 (0.1)	77 (0.1)	13 (0.1)
Sequence of radiation and surgery performed (n, %)			
No radiation and/or cancer-directed surgery	168 694 (52.3)	32 906 (55.6)	11 505 (57.3)
Radiation after surgery	149 615 (46.4)	25 718 (43.4)	8340 (41.5)
Intraoperative radiation	1839 (0.6)	173 (0.3)	68 (0.3)
Radiation prior to surgery	728 (0.2)	168 (0.3)	80 (0.4)
Radiation before and after surgery	673 (0.2)	146 (0.2)	40 (0.2)
Intraoperative rad with other rad before/after surgery	560 (0.2)	53 (0.1)	28 (0.1)
Other	239 (0.1)	40 (0.1)	12 (0.1)

Continued

Table 2 Continued

Reason no cancer-directed surgery (n, %)			
Surgery performed	305 404 (94.7)	55 371 (93.5)	18 308 (91.2)
Not recommended	13 032 (4.0)	2952 (5.0)	1356 (6.8)
Recommended, unknown if performed	1899 (0.6)	568 (1.0)	217 (1.1)
Recommended but not performed, patient refused	1364 (0.4)	211 (0.4)	146 (0.7)
Not recommended, contraindicated due to other	333 (0.1)	53 (0.1)	22 (0.1)
conditions, autopsy only	285 (0.1)	44 (0.1)	22 (0.1)
Other	31 (0.0)	5 (0.0)	2 (0.0)

N/A, not available.

ethnicity-specific models, Cox PH showed the highest c-index score followed by GBT, SSVM and ST.

DISCUSSION

Accurately predicting the survival outcome of a patient is critical in determining treatment options and providing appropriate cancer care. The ML approaches provide a robust way of predicting health outcomes using large data points with complex feature interactions. However, current ML models are often built with all races/ethnicity data, having the potential to have representation bias, and not tailored to each minority group. To date, race/ethnicity-specific survival ML models predicting the outcomes of the black and Hispanic women diagnosed with breast cancer are lacking. This study developed and evaluated race/ethnicity-specific survival ML models for black and Hispanic women with breast cancer and compared with the general population model. The high performing ML models developed in this study will be able to contribute to providing individualised oncology care and improving the survival outcome of specific populations, the black and Hispanic women. Also, it is a strength of our model that we used the patient data from more than 322 348 women in a large, population-based dataset from 2000 to 2017, including 59 204 (18.4%) Hispanic women and 20 073 (6.2%) Black women.

The sample population in this study showed that the black population had the highest death rate followed by the Hispanic and all races/ethnicity, supporting the findings from other literature.[4 5] Also, the survival months for the black and Hispanic groups were low and they were younger compared with all races/ethnicity. It is congruent with the literature that young black women have higher breast cancer mortality than young white women,[16 17] and the Latinas have the higher rates of more advanced cancer than non-Hispanic Whites.[18] Also, breast cancer is more aggressive in younger women than older premenopausal women.[19] Our study sample also showed that the Hispanic and black populations had higher percentage of poorly differentiated grade III cancer than overall populations. Poorly differentiated tumours lack normal features, tend to grow and spread faster and have a worse prognosis[20]; and these tumours expressed lower levels of oestrogen receptor.[21] Our study sample showed likewise that Hispanic and black populations showed the lower percentage of oestrogen receptor positive status and progesterone receptor positive status than overall population. Studies have shown that young age breast cancer has more advanced stage at presentation, more grades and higher oestrogen receptor negativity.[22]

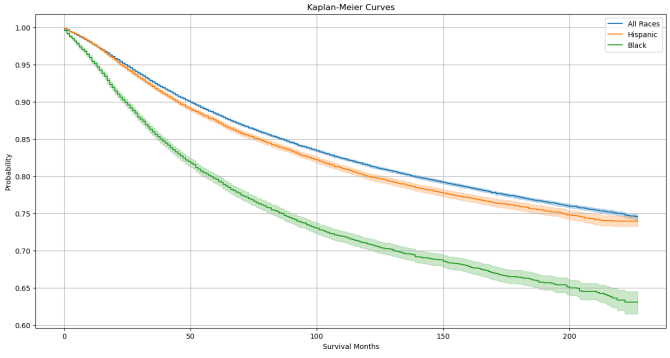

Figure 3 Kalplan-Meier survival curves.

The result also showed that lower percentages of Hispanic and black populations had chemotherapy. Existing literature has shown that African American and Hispanic patients tend to experience diagnostic and treatment delays, which were related to worse survival outcomes.[23 24] Perhaps lower percentages of Hispanic and black patients receiving chemotherapy were associated with the fewer survival months of the Hispanic and black populations in this study.

After the race/ethnicity-specific model development and evaluation, we observed that the general models trained with all races/ethnicity did not perform well when tested with specific races/ethnicity. That is, the race/ethnicity-specific survival ML models developed in this study consistently outperformed the general models when predicting the outcomes of specific race/ethnicity, addressing bias in ML. Especially, black and Hispanic-specific survival ML models using the Cox PH approach showed the best performance among the four ML models tested, showing that this model outperformed the other models in predicting the survival of specific race/ethnicity. Also, the ST model performance showed the

Table 3 Model performance comparison using c-index

	Hispanic-specific model	All races/ethnicity model	Black-specific model	All races/ethnicity model
Model	M_h	$M_{all,h}$	M_b	$M_{all,b}$
Test	E_h	E_h	E_b	E_b
Cox PH	0.832	0.828	0.823	0.821
GBT	0.813	0.813	0.808	0.803
ST	0.772	0.763	0.804	0.801
SSVM	0.834	0.790	0.824	0.786

GBT, Gradient Boost Tree; PH, proportional hazard; SSVM, survival support vector machine; ST, survival tree.

highest difference between the race/ethnicity-specific model and the general model. This indicates that the ST model tends to overfit to a specific race/ethnicity compared with the other models. Our study demonstrated that a tailored ML model for each race/ethnicity is needed to better predict the patient survival than the general ML model using all races/ethnicity. By accurately forecasting a patient's survival, healthcare professionals will be able to guide individualised treatment decisions and provide tailored interventions for the well-being of a cancer survivor.

It is worth noting that although the performance of the general model is not low, it was trained with the general population with an imbalanced portion of the underrepresented population, including the Hispanic and black populations. It was still meaningful to examine the feasibility of race/ethnicity-specific models since it is recommended to train an ML model with data resembling the people the model is intended to use to mitigate representation bias. Although the performance difference between the models was sometimes marginal depending on the algorithms, our race/ethnicity-specific models consistently outperformed the general model. It shows the potential to accurately predict individualised patient outcomes for quality care delivery for underrepresented populations and lead to alleviating health disparities.

There are several limitations to this study. The SEER database only includes the first course of treatment and do not have information on adjuvant therapy.[25] This causes difficulties comparing the outcomes of the treatment sequence. To overcome this limitation, a comprehensive database that has more information on cancer treatment can be used as a future work to provide additional insights on the impact of treatment sequence. Also, the dataset did not include the human epidermal growth factor 2 receptor status, which is a critical tumour marker for breast cancer prognosis. The variable was missing because it was collected from 2010, but our data were dated from 2000. Incorporating this variable in the modelling will be needed in future work to provide more accurate predictions for patient outcomes.

CONCLUSION

This study has developed and evaluated accurate race/ethnicity-specific survival ML models for black and Hispanic women diagnosed with breast cancer. Predicting the individualised survival outcome of breast cancer can

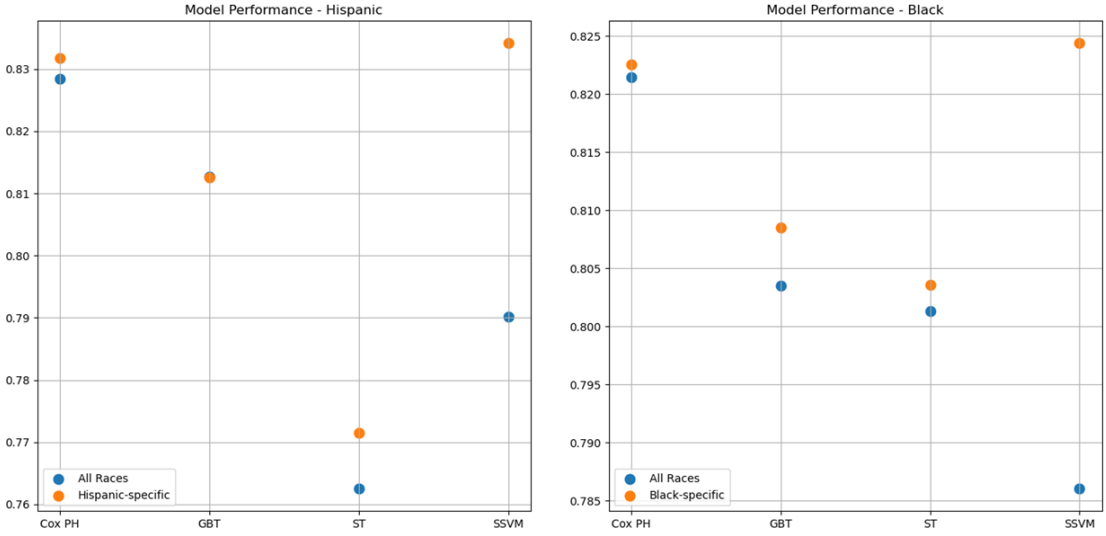

Figure 4 Race/ethnicity-specific model performance comparison using C-index. GBT, Gradient Boost Tree; PH, proportional hazards; SSVM, survival support vector machine; ST, survival tree.

provide the evidence necessary for determining treatment options and high-quality, patient-centred cancer care delivery for underrepresented populations. Also, the race/ethnicity-specific ML models can mitigate representation bias and contribute to addressing health disparities.

Contributors All the authors contributed to the design of the work and the final approval of the submission. JIP worked on the data acquisition, analysis and interpretation of the data, and acted as guarantor. SB contributed to the data analysis. JWP and SL contributed to the interpretation of data for the work.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Since the data were fully deidentified, this study was not considered human subject research by the Institutional Review Board at the University, and no informed consent was required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. We used the National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) Program data. It provides information on cancer statistics in an effort to reduce the cancer burden among the US population.

ORCID iD

Jung In Park http://orcid.org/0000-0002-1771-7361

REFERENCES

1 Siegel RL, Miller KD, Fuchs HE, *et al*. Cancer statistics, 2021. *CA Cancer J Clin* 2021;71:7–33.
2 Yedjou CG, Sims JN, Miele L. Health and racial disparity in breast cancer. *Breast cancer metastasis and drug resistance* 2019;1152:31–49.
3 Power EJ, Chin ML, Haq MM. Breast cancer incidence and risk reduction in the Hispanic population. *Cureus* 2018;10:e2235.
4 Jemal A, Ward EM, Johnson CJ, *et al*. Annual report to the nation on the status of cancer, 1975–2014, featuring survival. *J Natl Cancer Inst* 2017;109.
5 Copeland G, Green D, Firth R. Cancer in North America: 2011–2015 volume one: combined cancer incidence for the United States, Canada and North America. Springfield North American Association of Central Cancer Registries; 2018.
6 Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med* 2019;380:1347–58.
7 Bates DW, Saria S, Ohno-Machado L, *et al*. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff* 2014;33:1123–31.
8 Obermeyer Z, Powers B, Vogeli C, *et al*. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366:447–53.
9 Saria S, Subbaswamy A. Tutorial: safe and reliable machine learning. *arXiv preprint* 2019.
10 Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. *PLoS Med* 2018;15:e1002689.
11 Rajkomar A, Hardt M, Howell MD, *et al*. Ensuring fairness in machine learning to advance health equity. *Ann Intern Med* 2018;169:866–72.
12 Wiens J, Saria S, Sendak M, *et al*. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med* 2019;25:1337–40.
13 SEER, National Cancer Institute. SEER research plus data description cases diagnosed in 1975-2017, 2020. Available: https://seer.cancer.gov/data-software/documentation/seerstat/nov2019/TextData.FileDescription.pdf
14 Duggan MA, Anderson WF, Altekruse S, *et al*. The surveillance, epidemiology, and end results (SEER) program and pathology: toward strengthening the critical relationship. *Am J Surg Pathol* 2016;40:e94.
15 Chen Y, Jia Z, Mercola D, *et al*. A gradient boosting algorithm for survival analysis via direct optimization of concordance index. *Comput Math Methods Med* 2013;2013:1–8.
16 Shavers VL, Harlan LC, Stevens JL. Racial/Ethnic variation in clinical presentation, treatment, and survival among breast cancer patients under age 35. *Cancer* 2003;97:134–47.
17 Althuis MD, Brogan DD, Coates RJ, *et al*. Breast cancers among very young premenopausal women (United States). *Cancer Causes Control* 2003;14:151–60.
18 Yanez B, Thompson EH, Stanton AL. Quality of life among Latina breast cancer patients: a systematic review of the literature. *J Cancer Surviv* 2011;5:191–207.
19 Gonzalez-Angulo AM, Broglio K, Kau S-W, *et al*. Women age < or = 35 years with primary breast carcinoma: disease features at presentation. *Cancer* 2005;103:2466–72.
20 American Cancer Society. Understanding your pathology report: breast cancer, 2020. Available: https://www.cancer.org/treatment/understanding-your-diagnosis/tests/understanding-your-pathology-report/breast-pathology/breast-cancer-pathology.html
21 Scimeca M, Antonacci C, Colombo D, *et al*. Emerging prognostic markers related to mesenchymal characteristics of poorly differentiated breast cancers. *Tumour Biol* 2016;37:5427–35.
22 Zabicki K, Colbert JA, Dominguez FJ, *et al*. Breast cancer diagnosis in women < or = 40 versus 50 to 60 years: increasing size and stage disparity compared with older women over time. *Ann Surg Oncol* 2006;13:1072–7.
23 Gwyn K, Bondy ML, Cohen DS, *et al*. Racial differences in diagnosis, treatment, and clinical delays in a population-based study of patients with newly diagnosed breast carcinoma. *Cancer* 2004;100:1595–604.
24 Fedewa SA, Ward EM, Stewart AK, *et al*. Delays in adjuvant chemotherapy treatment among patients with breast cancer are more likely in African American and Hispanic populations: a national cohort study 2004-2006. *J Clin Oncol* 2010;28:4135–41.
25 Yu JB, Gross CP, Wilson LD, *et al*. NCI SEER public-use data: applications and limitations in oncology research. *Oncology* 2009;23:288.
26 Katzman JL, Shaham U, Cloninger A, *et al*. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol* 2018;18:1–2.
27 Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *J Am Stat Assoc* 1989;84:1074–8.
28 Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal* 2002;38:367–78.
29 Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann. Statist.* 2001;29:1189–232.
30 Bertsimas D, Dunn J, Gibson E, *et al*. Optimal survival trees. *Mach Learn* 2022;111:2951–3023.
31 Leblanc M, Crowley J. Survival trees by Goodness of split. *J Am Stat Assoc* 1993;88:457–67.
32 Van Belle V, Pelckmans K, Van Huffel S, *et al*. Support vector methods for survival analysis: a comparison between ranking and regression approaches. *Artif Intell Med* 2011;53:107–18.
33 Pölsterl S, Navab N, Katouzian A. An efficient training algorithm for kernel survival support vector machines. *arXiv preprint* 2016.

BMJ Health & Care Informatics

Effect of a hospital command centre on patient safety: an interrupted time series study

Teumzghi F Mebrahtu ,[1,2] Ciarán D McInerney ,[1,3] Jonathan Benn,[2,4] Carolyn McCrorie,[3,4] Josh Granger,[4] Tom Lawton,[5] Naeem Sheikh,[3] Rebecca Randell,[6,7] Ibrahim Habli,[8] Owen Ashby Johnson[1,3]

To cite: Mebrahtu TF, McInerney CD, Benn J, *et al*. Effect of a hospital command centre on patient safety: an interrupted time series study. *BMJ Health Care Inform* 2023;**30**:e100653. doi:10.1136/bmjhci-2022-100653

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjhci-2022-100653).

Received 22 July 2022
Accepted 19 December 2022

For numbered affiliations see end of article.

Correspondence to
Dr Teumzghi F Mebrahtu; teumzghi.mebrahtu@bthft.nhs.uk

ABSTRACT

Background Command centres have been piloted in some hospitals across the developed world in the last few years. Their impact on patient safety, however, has not been systematically studied. Hence, we aimed to investigate this.

Methods This is a retrospective population-based cohort study. Participants were patients who visited Bradford Royal Infirmary Hospital and Calderdale & Huddersfield hospitals between 1 January 2018 and 31 August 2021. A five-phase, interrupted time series, linear regression analysis was used.

Results After introduction of a Command Centre, while mortality and readmissions marginally improved, there was no statistically significant impact on postoperative sepsis. In the intervention hospital, when compared with the preintervention period, mortality decreased by 1.4% (95% CI 0.8% to 1.9%), 1.5% (95% CI 0.9% to 2.1%), 1.3% (95% CI 0.7% to 1.8%) and 2.5% (95% CI 1.7% to 3.4%) during successive phases of the command centre programme, including roll-in and activation of the technology and preparatory quality improvement work. However, in the control site, compared with the baseline, the weekly mortality also decreased by 2.0% (95% CI 0.9 to 3.1), 2.3% (95% CI 1.1 to 3.5), 1.3% (95% CI 0.2 to 2.4), 3.1% (95% CI 1.4 to 4.8) for the respective intervention phases. No impact on any of the indicators was observed when only the software technology part of the Command Centre was considered.

Conclusion Implementation of a hospital Command Centre may have a marginal positive impact on patient safety when implemented as part of a broader hospital-wide improvement programme including colocation of operations and clinical leads in a central location. However, improvement in patient safety indicators was also observed for a comparable period in the control site. Further evaluative research into the impact of hospital command centres on a broader range of patient safety and other outcomes is warranted.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Although command centres have been introduced in hospitals in developed world countries, their impact on patient safety remains unknown.

WHAT THIS STUDY ADDS

⇒ A command centre that includes introduction of both technological (data display) elements and organisational components may improve patient safety. However, it appears that the majority of the impact may result from the processes around the command centre itself rather than the technological aspect.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ It is a common belief that command centres can improve patient safety. However, this has not been supported by patient safety metrics examined for this study. Hence, further research is warranted.

INTRODUCTION

Fragmented healthcare is neither cost-effective nor safe for the delivery of patient care.[1 2] In most UK National Health Service (NHS) hospitals, health service delivery is fragmented across multiple departments and services with major implications for patient safety, efficiency and good patient care. Such fragmentation can, however, be minimised using health information technology to improve the flow of information—between and within healthcare providers.[3 4] The idea of improving communication by using digital information systems to centralise information to improve situational awareness was pioneered by the National Aeronautics and Space Administration (NASA) for the purpose managing space flights six decades ago.[5] This central system, also known as 'command centre' or 'mission control' has been widely adopted in retail industries, finance and banking, automotive, manufacturing and transport industries and to a lesser degree within the healthcare sector.

In the last 5 years, a number of hospitals in Canada, China, the UK, USA and Saudi Arabia have been piloting 'command centres' for the purpose of patient-flow management. Although not from systematically conducted studies, preliminary reports suggest that

command centres have a positive impact on patient care delivery process.[6–10] For example, in Johns Hopkins Hospital USA, patient transfers from other hospitals improved by 46%, ambulances dispatches reduced by 43 min and bed allocation for emergency admission patients reduced by 3.5 hours.[7]

In the UK, there are currently only four NHS hospital trusts who are piloting command centres. One of these is Bradford Teaching Hospitals NHS Foundation Trust, which provides hospital services for around half a million people. In 2019, the Trust introduced a command centre at its main hospital, Bradford Royal Infirmary (BRI).[11] The command centre is made up of software and display screens (also known as 'tiles') that provides real-time information (updated every 3 min) and alerts for patient care and intervention across the hospital site, including: overall hospital capacity, emergency department status, patient transfers, discharge tasks, care progression and patient deterioration. Information is inputted by the staff in the departments of the BRI hospital as part of normal care processes within the electronic patient record system and is automatically reconfigured to be shown in defined parameters within each of the tiles.

The Bradford Command Centre aims to provide safer care by addressing increasing pressure in the ED and associated challenges downstream related to capacity and demand, monitoring patients for placement in most appropriate care settings and access to real-time information required to make decisions. Such command centres have the potential to improve future patient flow and safety, and research to understand the health service delivery, safety and operational factors is considered an area of major importance for hospitals. We hypothesised that the implementation of an integrated and centralised hospital command centre improves patient safety. Therefore, our study aim was twofold: (1) to investigate the impact of Bradford command centre on patient safety outcomes in BRI hospital (2) to compare the pattern of patient safety outcomes of BRI hospital with Calderdale & Huddersfield Hospitals (CHH) which is without a command centre.

METHODS

Study population

Participants of the study were patients who visited accident and emergency and unplanned admissions at the intervention site, BRI hospital Trust, where the command centre was introduced and a nearby and similar sised hospital, CHH, which we used as a control. The study period was between 1 January 2018 and 31 August 2021 and covered the period before, during and after the implementation of the command centre at the intervention site.

Study design

This is a retrospective population-based cohort study undertaken as part of a mixed method evaluation project with a formal evaluation protocol published by the authors in January 2022.[12] Qualitative study of the command centre programme gave rise to two hypothesised intervention timelines, one focusing on the implementation and activation of the technological components of the command centre and the other 'complex' intervention model that sought to account for the broader patient flow and operational redesign programme in which the command centre technology was a part. For the technology model, a three-phase, interrupted time series model was used to reflect incremental implementation of the visual displays in the command centre, consisting of a preintervention (baseline), first intervention component ('command centre displays roll-in') and second intervention component ('command centre activation'). For the complex intervention model, a five-phase, interrupted time series model was used that consisted of preintervention (baseline), first intervention component ('onset of patient flow programme'), second intervention component ('command centre displays roll-in'), third intervention component ('command centre activation') and fourth intervention component ('hospital wide engagement and training'), the latter referring to roll-out of remote access to command centre data across the hospital. See table 1 for the details of the timeline and interrupts.

Table 1 Project timeline and intervention phases

Date	Event
1 January 2018	Start of study
1 July 2018	Onset of patient flow programme
1 May 2019	Command centre displays roll-in
1 December 2019	Command centre activation and hospital wide engagement and training commences
1 May 2021	Post-COVID-19 resumption of hospital wide engagement and training
31 August 2021	End of study

Data source

UK NHS Digital Secondary Use Services (SUS) data were used. These data are secure, patient-level data that is sent by both hospitals to NHS England to support national tariff policy and secondary analysis. Construction of the SUS data was conducted by Connected Bradford, a team located at the Bradford Institute for Health Research (https://www.bradfordresearch.nhs.uk/our-research-teams/connected-bradford/). The team uploaded the SUS data onto a Google Cloud Platform where relevant data were processed before final outputs were extracted.[13]

Patient and public involvement and engagement

Public and patient representatives contributed to the development of the research protocol and towards selection of proxy patient safety outcomes. Details of the wider

project's patient and public involvement and engagement are available in the published protocol.[12]

Outcome variables

Potential patient safety outcome indicators were listed in our published protocol.[12] However, due to data availability, three indicators for the BRI hospital (mortality, readmissions within 72 hours and postoperative sepsis) and two indicators for the CHH (mortality and readmissions within 72 hours) were analysed.

The proportions of readmissions and mortality in hospital were calculated as the total readmissions and death among emergency admissions, respectively, divided by the total number of emergency admissions. Postoperative sepsis was calculated by dividing the weekly count of patients with sepsis diagnostic codes in their records by the count of surgical operations conducted in that week. The list of surgical operation codes was extracted from the UK Health Security Agency published document.[14] Postoperative sepsis occurrences were identified using T814 ICD10 code.[15]

Variables for analysis

Dummy variables were created for each of the intervention components ('Onset of patient flow programme', 'command centre displays roll-in', 'command centre activation' and 'hospital wide engagement and training'), COVID-19 pandemic and spikes of COVID-19 pandemic.[16] The components of the intervention were given a value of '1' starting from the date of its introduction until the introduction of the next component or phase, then a value of '0' for the rest of the period. 'COVID-19 pandemic' was given a value of '0' through February 2020 and a value of '1', thereafter. A spike dummy variable was also added by setting '1' for the COVID-19 spike periods based on the UK data[16] and '0' throughout.

A continuous incremental time variable was coded from the start of the time series (eg, 1, 2, 3, 4). The intervention phases were also modelled using five continuous time variables with '0' in the preintervention period, '1, 2, 3, 4….' from the onset of the intervention phase. In addition, seasonality was modelled by including dummy variables for the number of weeks in the year.

Statistical analysis and software

First, outcome variables were summarised descriptively. Then, to assess the impact of the command centre on patient safety outcomes, linear regression interrupted time series analysis was used.[17] Linear time series models were fitted to the BRI and CHH data separately. Tests for serial autocorrelation of residuals were conducted and all tests were non-statistically significant. Hence, regression models with autoregressive integrated moving average (ARIMA) errors were not used. To compare the changes (ie, beta coefficients) of two outcomes (mortality and readmissions within 72 hours) in BRI and CHH, the difference of the changes (beta coefficient BRI subtracted beta coefficient CHH) and total variances (variance BRI+variance CHH) were calculated to derive the difference in changes and their CIs.[18]

Akaike information criterion[19] and Bayesian information criterion[20] were used to select the best fitting models. Analyses were implemented in R (V.4.0.2). We adopted 5% significance levels and 95% CIs throughout.

A five-phase interrupted time series was used for the main analyses. To explore if the command centre activation alone would have had an impact on outcomes, a three-phase interrupted time series model was used for sensitivity analyses for the intervention site (BRI hospital), including the display roll-in period and activation event.

RESULTS

Descriptive summary

There were a total of 203 807 in-patient emergency admissions and 34 625 operations performed in BRI hospital and 291 018 in-patient emergency admissions in CHH during the study period. The weekly mortality (as a percentage of weekly admissions) stayed below 3% and 5% (in BRI and CHH, respectively) for most of the study period except a sudden increase in March–April 2020 when a spike in the hospital admissions associated with COVID-19 was reported in the UK.[16] The weekly deaths appear to be higher for the period after COVID-19 pandemic when compared with the prepandemic period. Overall, the average mortality throughout the study period was higher in CHH than BRI (see figure 1 and table 2).

The weekly readmissions within 72 hours (as a percentage of the total emergency admissions) remained above 6% and 4% (in BRI and CHH, respectively) for the majority of the study period. The average readmissions in BRI hospital were just over 8% during the first 6 months of the study period and stayed just under 7% for the remaining period. On the other hand, the readmissions in the CHH were under 3% during the first 16 months, then nearly doubled during the rest of the study period. The patterns of the weekly readmissions do not appear to have been greatly affected by the pandemic. Overall, the weekly readmissions were higher in BRI than CHH (see figure 1 and table 2).

The weekly postoperative sepsis (as a percentage all surgical operations performed) stayed between 1.5% and 6% for majority of the period with occurrence of spikes during January and April 2020 (see figure 1). The overall postoperative sepsis ranged between 0.6% and 10%, and it was below 5% during the study period on average (see figure 1 and table 2).

The effect of intervention

Main analyses (five-phase interrupted time series)

In BRI hospital, when compared with the preintervention period, the weekly mortality decreased by 1.4% (95% CI 0.8% to 1.9%), 1.5% (95% CI 0.9% to 2.1%), 1.3% (95% CI 0.8% to 1.9%) and 2.5% (95% CI 1.7% to 3.4%) at onset of the first ('patient flow programme'), second ('command centre display roll-in'), third ('command

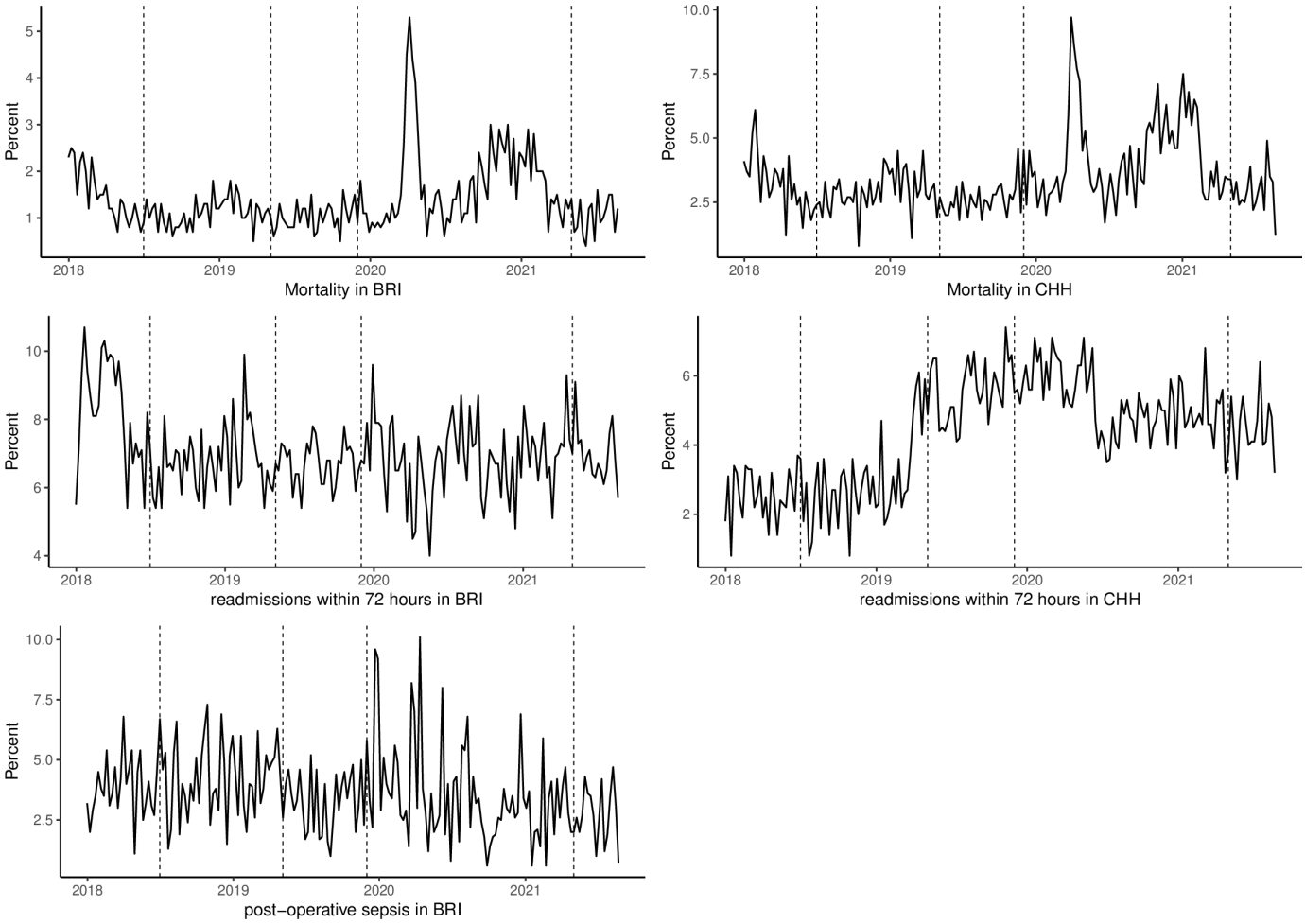

Figure 1 An overall pattern of patient safety indicators during the study period. BRI, Bradford Royal Infirmary; CHH, Calderdale & Huddersfield Hospitals.

centre activation') and fourth ('hospital-wide engagement resumption') intervention periods, respectively. At onset of the first, second and third intervention periods, the weekly per cent of readmission within 72 hours also decreased by 2.7% (95% CI 1.7% to 3.8%), 2.5% (95% CI 1.4% to 3.6%), 2.0% (95% CI 1.0% to 3.0%) and 0.7% (95% CI 2.2% to 0.9%), respectively. The weekly postoperative sepsis did not show a significant change

Table 2 Summary of patient safety indicators, mean (SD)

	Mortality (%)*		Readmissions within 72 hours (%)*		Post-operative sepsis (%)†
Period	BRI	CHH	BRI	CHH	BRI
1 January 2018–30 June 2018 (pre-intervention)	1.5 (0.56)	3.3 (1.1)	8.2 (1.5)	2.5 (0.7)	3.8 (1.2)
1 July 2018–30 April 2019 (patient flow programme)	1.1 (0.31)	2.9 (0.78)	6.8 (0.97)	2.9 (1.22)	4.2 (1.6)
1 May 2019–30 November 2019 (command centre display roll-in)	1.0 (0.29)	2.5 (0.57)	6.7 (0.62)	5.6 (0.87)	3.2 (1.2)
1 December 2019–30 April 2021 (command centre activation)	1.7 (0.94)	4.3 (1.67)	6.8 (1.1)	5.2 (0.93)	3.6 (2.1)
1 May 2021–31 August 2021 (engagement resumption)	1.1 (0.38)	2.9 (0.83)	6.9 (0.81)	4.5 (0.84)	2.7 (1.2)

*Values are percentages with respect to weekly counts of in-patient emergency admissions.
†Values are percentages with respect to weekly counts of surgical operations.
BRI, Bradford Royal Infirmary hospital; CHH, Calderdale & Huddersfield Hospitals.

Table 3 Summary results for five-phase models

Outcome	Intervention phase	Change in BRI (95% CI)*	Change in CHH (95% CI)*	Difference between sites (BRI–CHH), 95% CI †
Mortality (%)	Pre-intervention	Ref.	Ref.	
	Patient flow programme	−1.4 (−1.9 to −0.8)	−2.0 (−3.1 to −0.9)	0.6 (−0.6 to 1.9)
	Command centre display roll-in	−1.5 (−2.1 to −0.9)	−2.3 (−3.5 to −1.1)	0.8 (−0.6 to 2.2)
	Command centre activation	−1.3 (−1.82 to −0.7)	−1.3 (−2.4 to −0.2)	0.04 (−1.2 to 1.3)
	Engagement resumption	−2.5 (−3.4 to −1.7)	−3.1 (−4.8 to −1.4)	0.6 (−1.4 to 2.5)
Readmissions within 72 hours (%)	Pre-intervention	Ref.	Ref.	
	Patient flow programme	−2.7 (−3.8 to −1.7)	−0.6 (−1.5 to 0.2)	−2.1 (−3.4 to −0.7)
	Command centre display roll-in	−2.5 (−3.6 to −1.4)	2.6 (1.6 to 3.5)	−5.1 (−6.6 to −3.6)
	Command centre activation	−2.02 (−3.0 to −1.0)	3.6 (2.7 to 4.5)	−5.6 (−6.9 to −4.3)
	Engagement resumption	−0.70 (−2.3 to 0.9)	2.2 (0.8 to 3.5)	−2.9 (−4.8 to −0.8)
Post-operative sepsis (%)	Pre-intervention	Ref.	–	–
	Patient flow programme	0.4 (−1.2 to 2.0)	–	–
	Command centre display roll-in	−0.5 (−2.2 to 1.3)	–	–
	Command centre activation	1.31 (−0.3 to 2.9)	–	–
	Engagement resumption	−0.2 (−2.7 to 2.2)	–	–

*Models were adjusted for trend, COVID-19 pandemic (pre-pandemic and post-pandemic) and COVID-19 spikes.
†Calculated as difference in changes (change BRI–change CHH) and total variances (variance BRI+variance CHH).
BRI, Bradford Royal Infirmary hospital; CHH, Calderdale & Huddersfield Hospitals.

during the study period, however (see table 3 and figure 2).

In CHH, compared with the baseline, the weekly mortality decreased by 2.0% (95% CI 3.1 to 0.9), 2.3% (95% CI 3.5 to 1.1), 1.3% (95% CI 2.4 to 0.2), 3.1% (95% CI 4.8 to 1.4) for the respective intervention phases of BRI hospital. However, except for the first intervention period, readmissions within 72 hours showed a significant increase during the second (change=2.6%, 95% CI 1.6 to 3.5), third (change=3.6, 95% CI 2.7 to 4.5) and forth (change=2.2, 95% CI 0.8 to 3.5), see table 3 and online supplemental table 1.

When the BRI hospital and CHH are compared in terms of indicator outcome changes during the study period, the weekly mortality significantly improved while the weekly readmissions showed improvement in BRI hospital but not in CHH (see table 3).

Sensitivity analysis (three-phase interrupted time series)

When implementation and activation of the technological aspects of the command centre were modelled, there was no significant difference between the pre and postintervention periods in the patient safety indicators. For example, mortality did not significantly change after the 'command centre display roll-in' (change=−0.5%, 95% CI −1.3 to 0.3) and 'command centre activation' (change=−0.3, 95% CI −1.0 to 0.4) periods when compared with the preintervention period (see online supplemental table 2).

DISCUSSION

In this preintervention and postintervention comparative study using SUS data, the findings indicate that introduction of the Bradford Command Centre may have improved patient safety. However, given improvements in mortality have also been observed in the CHH (control site) during the same period, improvements seen in the BRI hospital data may not be entirely due to the command centre. In addition, there was no significant difference between preintervention and postintervention periods linked to only the technological components of the command centre system.

Hospital command centres are expected to improve the management of patient flow by making use of real-time monitoring of patients. It is hypothesised that this improved patient flow is beneficial for patient safety. However, in our related work,[21] we found that measures of patient flow did not indicate improvements at the BRI site during the study period. This suggests that patient flow may not be primarily responsible for the improvements. Given also that similar improvements were seen at our control site, it could be that the changes observed by our measures of patient safety were due to nationwide responses to the COVID-19 pandemic or some other within-hospital factors that we did not measure. The impact of command centres on patient safety in complex multiple department hospitals is rarely reported in the literature, mainly due to the novelty of this type of initiative in acute care. A recent report from Saudi National Health Command Centre (NHCC) indicated

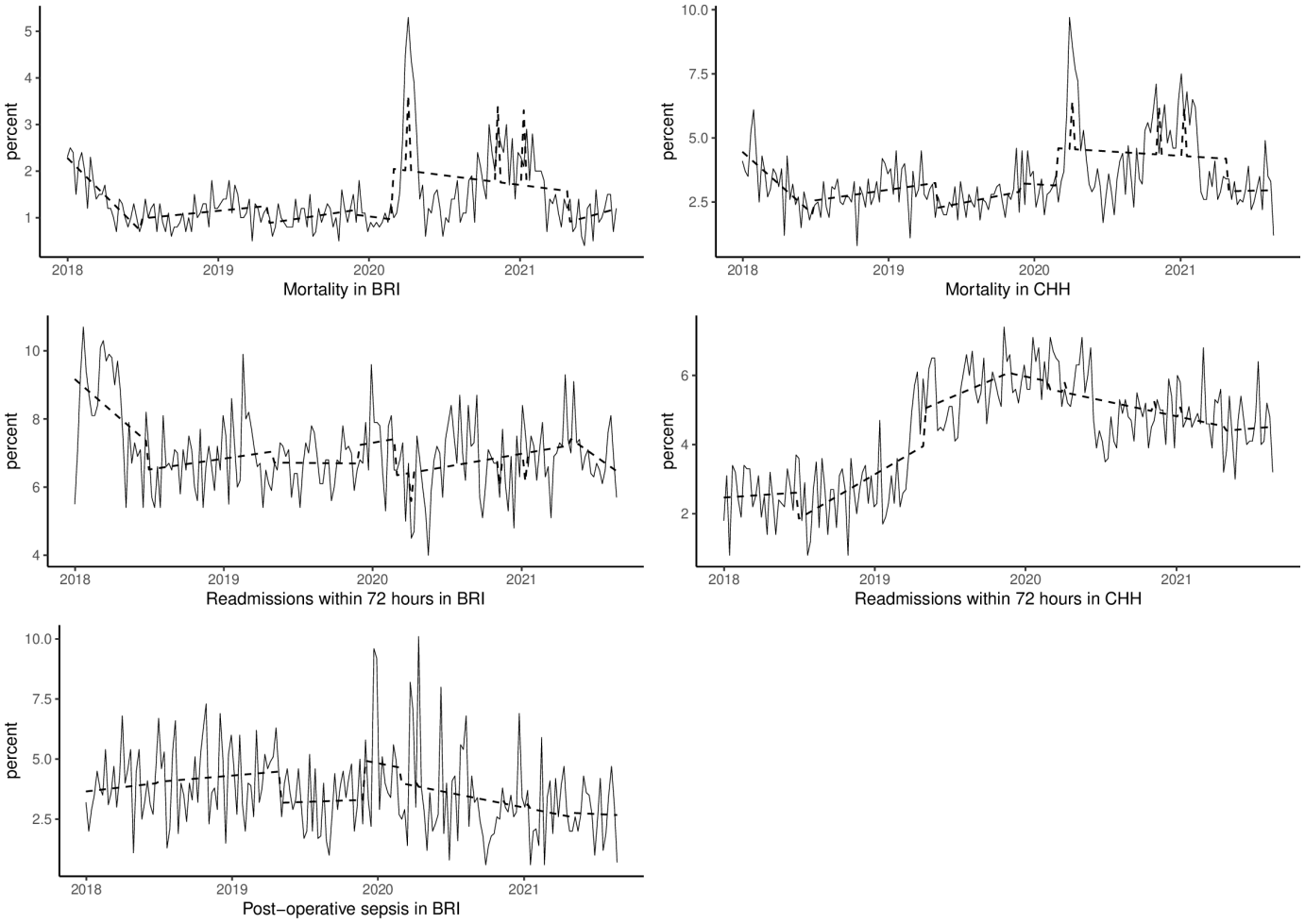

Figure 2 Actual values (solid lines) and model estimated values (dashed lines) of outcome indicators. BRI, Bradford Royal Infirmary; CHH, Calderdale & Huddersfield Hospitals.

that emergency admissions mortality was below 2%.[6] The mortality rate reported by the authors agrees with our findings of this study. What must be noted though is that the NHCC is a hub at a national level and its report appears to have compared pre-COVID-19 and post-COVID-19 pandemic data. On the other hand, the Bradford command centre is a single trust hospital and our study has compared preintervention period data against the multistage postintervention data, which included the prepandemic and postpandemic period.

Our study has certain limitations. First, health service delivery was significantly affected by the COVID-19 pandemic resulting in rapid systemwide effects, which may have impacted on the population of patients and capacity management in both hospitals. Cancellation and postponement of surgical operations were common due to reallocation of resources during the peaks of the pandemic. Although we attempted to control for the effects of the pandemic in our time series models, the proximity of the activation of the command centre with the onset of the pandemic surge makes it difficult to isolate the effect of the intervention or control for the pandemic without masking potential variation.

Second, apart from the command centre, it has been assumed that the intervention site (BRI hospital) and control site (CHH) are equivalent in other factors, which may not necessarily be the case. The control site showed considerably higher initial mortality which might have led to subsequent reduction in mortality rates or local interventions to reduce mortality, acting as a confounding factor in attempts to isolate the effect of the command centre intervention. Readmission rates additionally showed widely different trends between the study and control site.

Another potential limitation of the study concerns the focus of this quantitative evaluation on a small number of outcome indicators for what was a system-wide initiative designed to impact many areas. Although informing our intervention models using qualitative research at the study site is a strength in our design, qualitative investigation additionally revealed the complexity of this type of intervention and the challenges of implementation within a pressured acute care environment. This may have influenced the study outcome in a number of ways. Staff recall of the historical implementation timeline was variable (especially for piloting and roll-in of intervention components,

including organisational in addition to technological elements). There were suggestions that colocation of staff in the command centre room preceded the roll-in and activation phase for command centre displays, so the team may have already been established and coordinating functions sooner than the intervention timeline suggests, leading to under specification of our model. When considering the challenges observed in implementing the technological aspects of the intervention, including data quality, there may have been significant time lag between activation of components and any impact on patient safety outcomes. Given the complexity in our intervention model, we did not seek to control for lagged effects of intervention implementation (the time it takes for an intervention to start to influence detectable outcomes). Rather, we presumed that the effects of the intervention components were instantaneous.

Finally, due to data access limitations, we were not able to explore all outcomes identified for analysis in our study protocol. Hence, evaluation is needed, across multiple healthcare systems and command centre models, to understand how this type of intervention impacts downstream patient safety outcomes.

Nonetheless, the strengths of the study are threefold. First, we have used a large sample size for the analyses: a total of inpatient 203 807 inpatient visits and 34 625 surgical operations. Second, the use of electronic health record data minimises the inherent biases and errors in other types of observational data. Third, we employed a robust quasi-experimental design using repeated time series measurement.

In conclusion, the results of the study indicate that a digital hospital command centre package that includes both technological (data display) elements and organisational components may have a marginal positive impact on some patient safety outcomes. However, patient safety improvements in the control site hospital suggest that it may not entirely be due to the introduction of the command centre. In addition, when the technology alone was considered as the intervention (command centre display roll-in and command centre activation), it does not appear to have a significant impact on patient safety outcomes. Thus, further research using data from other hospital organisations that use command centres is warranted.

Author affiliations
[1]School of Computing, University of Leeds, Leeds, UK
[2]Bradford Institute for Health Research, Bradford, UK
[3]Wolfson Centre for Applied Health Research, Bradford Royal Infirmary, Yorkshire and Humber Patient Safety Translational Research Centre, Bradford, UK
[4]School of Psychology, University of Leeds, Leeds, UK
[5]Bradford Royal Infirmary, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK
[6]Faculty of Health Studies, University of Bradford, Bradford, UK
[7]Bradford Royal Infirmary, Wolfson Centre for Applied Health Research, Bradford, UK
[8]Department of Computer Science, University of York, York, UK

Twitter Ciarán D McInerney @CMc_PhD

Acknowledgements We thank the PPIE representatives for their contribution to the conception, development, ongoing implementation and future communication of this project. This study is based on data from Connected Bradford (REC 18/YH/0200 & 22/EM/0127). The data are provided by the citizens of Bradford and district and collected by the NHS and other organisations as part of their care and support.

Contributors JB, CMC, CDM and OAJ conceptualised the study. TFM, JB, CMC, CMI, JG, NS and OAJ contributed to the study design. TFM conducted data extraction, management, analysis and preparation of first draft of the manuscript. JB, CMC, CDM, JG, NS, IH, TL, RR and OAJ reviewed the manuscript for important intellectual content. TFM is guarantor.

Funding This research is supported by the National Institute for Health Research (NIHR) Yorkshire and Humber Patient Safety Translational Research Centre (NIHR Yorkshire and Humber PSTRC).

Disclaimer The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval The study was approved by the NHS Health Research Authority (IRAS Number: 285933) and the University of Leeds Engineering and Physical Sciences Research Ethics Committee (#MEEC 20-016).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iDs
Teumzghi F Mebrahtu http://orcid.org/0000-0003-4821-2304
Ciarán D McInerney http://orcid.org/0000-0001-7620-7110

REFERENCES

1 Stange KC. The problem of fragmentation and the need for integrative solutions. *Ann Fam Med* 2009;7:100–3.
2 Castro-Sánchez E, Charani E, Drumright LN, *et al*. Fragmentation of care threatens patient safety in peripheral vascular catheter management in acute care -- a qualitative study. *PLoS One* 2014;9:e86167
3 Nguyen L, Bellucci E, Nguyen LT. Electronic health records implementation: an evaluation of information system impact and contingency factors. *Int J Med Inform* 2014;83:779–96.
4 Lau F, Bartle-Clar J, Bliss G. Improving usability, safety and patient outcomes with health information technology. In: *From Research to Practice*. IOS Press, 2019.
5 National Aeronautics and Space Administration (NASA). Johnson space center's mission control center 2013. 2022. Available: https://www.nasa.gov/content/johnson-space-centers-mission-control-center-1/#.YgJk19_P2Uk [Accessed 8 Feb 2022].
6 Alharbi M, Senitan M, Ohanlon T, *et al*. 27 healthcare in the time of a pandemic and beyond: the innovative large-scale and integrated saudi national health command centre. *Leaders in Healthcare 2021* 2021;A10–1.
7 Hopkins J. Capacity command center celebrates 5 years of improving patient safety. 2021. Available: https://www.hopkinsmedicine.org/news/articles/capacity-command-center-celebrates-5-years-of-improving-patient-safety-access [Accessed 7 Feb 2022].
8 Kane EM, Scheulen JJ, Püttgen A, *et al*. Use of systems engineering to design a hospital command center. *Jt Comm J Qual Patient Saf* 2019;45:370–9.

9 Davenport PB, Carter KF, Echternach JM, *et al*. Integrating high-reliability principles to transform access and throughput by creating a centralized operations center. *J Nurs Adm* 2018;48:93–9.
10 Schlicher J, Metsker MT, Shah H, *et al*. FROM NASA TO HEALTHCARE: REAL-TIME DATA ANALYTICS (MISSION CONTROL) IS RESHAPING HEALTHCARE SERVICES. *Perspect Health Inf Manag* 2021;18:1g
11 Bradford Teaching Hospitals. Command centre. 2019. Available: https://www.bradfordhospitals.nhs.uk/command-centre/ [Accessed 8 Feb 2022].
12 McInerney C, McCrorie C, Benn J, *et al*. Evaluating the safety and patient impacts of an artificial intelligence command centre in acute hospital care: a mixed-methods protocol. *BMJ Open* 2022;12:e054090
13 Sohal K, Mason D, Birkinshaw J, *et al*. Connected Bradford: a whole system data linkage accelerator. *Wellcome Open Res* 2022;7:26.
14 UK HEALTH SECURITY AGENCY. *Updated operating procedure codes supplement (OPCS) and protocol for the surveillance of surgical site infection (derived from: NHS digital/health and social care information centre OPCS-4.8 to OPCS-4.9*). UK HEALTH SECURITY AGENCY, 2022: 1–88.
15 World Health Organisation (WHO). *ICD-10 version:2010*. WHO, 2010. Available: https://icd.who.int/browse10/2010/en#/T81.4
16 UK Health Security Agency. Coronavirus (COVID-19) in the UK 2022. Available: https://coronavirus.data.gov.uk/details/cases?areaType=overview&areaName=United%20Kingdom [Accessed 25 Feb 2022].
17 Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol* 2017;46:348–55.
18 Dias S, Caldwell DM. Network meta-analysis explained. *Arch Dis Child Fetal Neonatal Ed* 2019;104:F8–12.
19 Sakamoto Y, Ishiguro M, Kitagawa G. Akaike information criterion statistics. *Dordrecht Netherlands D Reidel* 1986;81:26853.
20 Schwarz G. Estimating the dimension of a model. *Ann Statist* 1978;6:461–4.
21 Mebrahtu TF, McInerney C, Benn J, *et al*. The impact of hospital command centre on patient flow and data quality: findings from the UK NHS. 2022;